

[illegible]

Db 2216 TCGGGCGGGATCGCGCTGCCGCTGCCGAGCTACCCCGACCGACCTCCGGGATCGG 2275
QY 1244 accgtggtatcgacgggtctgtagaacatggtccgcgaataccgacgacacacccgagcgc 1303
Db 2276 TGAAGCGGCTCTCGACGACCGCGGCCACCGCGCGCGCGCGGATGCGCGGACGACA 2335
QY 1304 tgcctcccgctggacacgactggtccgcggtccgcgagcggtgacgacggtggcaccgg 1360
Db 2336 TGTCTCGGAGCGCTACCGCGCGGAGGTCTCGGCATCTGCAGGACTGCGCGCGAG 2392

RESULT 7

US-07-642-734C-3
Sequence 3, Application US/07642734C
Patent No. 5824513

GENERAL INFORMATION:

APPLICANT: Katz, L
APPLICANT: Donadio, S
APPLICANT: Mcalpine, J B
TITLE OF INVENTION: Recombinant DNA Method for Producing
NUMBER OF INVENTION: Erythromycin Analogs
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:

ADDRESSEE: Edward H. Gorman
STREET: Abbott Laboratories D377/RP6D-2 One Abbott
CITY: Park Rd
STATE: IL

COUNTRY: US

ZIP: 60064-3500

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/642,734C

FILING DATE: 17-JAN-91

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Danckers, Andreas M

REGISTRATION NUMBER: 32652

REFERENCE/DOCKET NUMBER: 4952.US.01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 708-937-9396

TELEFAX: 708-938-2623

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 20235 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Saccharopolyspora erythraea

STRAIN: NRRL 238

FEATURE:

NAME/KEY: CDS

LOCATION: 19..10722

OTHER INFORMATION: /codon_start=19

OTHER INFORMATION: /function="gene eryA"

OTHER INFORMATION: /product="eryA ORF2 encoding modules 3 & 4 for

OTHER INFORMATION: 6-deoxyerythronolide B"

FEATURE:

NAME/KEY: misc_feature

LOCATION: 19..4470

OTHER INFORMATION: /function="approximate span of

OTHER INFORMATION: module 3"

FEATURE:

NAME/KEY: misc_feature

LOCATION: 97..1482
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..10722
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..5847
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function="approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start=10723
OTHER INFORMATION: /function="gene -eryA"
OTHER INFORMATION: /product="orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function="approximatr span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610

Search completed: September 26, 2000, 19:23:33
Job time: 12928 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2000, 15:47:32 ; Search time 5060.92 Seconds
(without alignments)
2055.201 Million cell updates/sec

Title: US-09-430-029-1
Perfect score: 5828
Sequence: 1 gatcattcatcaaatgcgc.....tcggccgggcaacacgacgac 5828

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 972840 seqs, 892348106 residues 1945680
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl:*
- 1: gb_ba1:*
 - 2: gb_ba2:*
 - 3: gb_om:*
 - 4: gb_ov:*
 - 5: gb_pat:*
 - 6: gb_ph:*
 - 7: gb_pl1:*
 - 8: gb_pl2:*
 - 9: gb_pr1:*
 - 10: gb_pr2:*
 - 11: gb_pr3:*
 - 12: gb_ro:*
 - 13: gb_sts:*
 - 14: gb_sy:*
 - 15: gb_un:*
 - 16: em_fun:*
 - 17: em_hum1:*
 - 18: em_hum2:*
 - 19: em_in:*
 - 20: em_om:*
 - 21: em_or:*
 - 22: em_ov:*
 - 23: em_pat:*
 - 24: em_ph:*
 - 25: em_pl:*
 - 26: em_ro:*
 - 27: em_sts:*
 - 28: em_sy:*
 - 29: em_un:*
 - 30: em_vi:*
 - 31: gb_htg1:*
 - 32: gb_htg2:*
 - 33: gb_in1:*
 - 34: gb_in2:*
 - 35: em_ba1:*
 - 36: em_ba2:*
 - 37: em_hum3:*
 - 38: em_hum4:*
 - 39: gb_pr4:*
 - 40: gb_htg3:*
 - 41: gb_htg4:*
 - 42: gb_htg5:*
 - 43: gb_htg6:*

- 44: gb_htg7:*
- 45: em_htg1:*
- 46: em_htg2:*
- 47: em_htg3:*
- 48: em_hum5:*
- 49: gb_pl3:*
- 50: gb_pr5:*
- 51: gb_htg8:*
- 52: gb_htg9:*
- 53: gb_htg10:*
- 54: gb_htg11:*
- 55: gb_htg12:*
- 56: gb_htg13:*
- 57: gb_htg14:*
- 58: gb_in3:*
- 59: gb_htg15:*
- 60: gb_htg16:*
- 61: gb_htg17:*
- 62: em_htg4:*
- 63: em_htg5:*
- 64: em_htg6:*
- 65: em_htg7:*
- 66: em_hum6:*
- 67: gb_htg18:*
- 68: gb_htg19:*
- 69: gb_htg20:*
- 70: gb_htg21:*
- 71: gb_htg22:*
- 72: gb_htg23:*
- 73: gb_htg24:*
- 74: gb_htg25:*
- 75: gb_htg26:*
- 76: gb_htg27:*
- 77: gb_htg28:*
- 78: gb_htg29:*
- 79: gb_htg30:*
- 80: gb_htg31:*
- 81: gb_vil:*
- 82: gb_vl2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		DB ID	Description
		Match	Length		
1	2738	47.0	6298	1 AB006479	AB006479 Comamonas
2	2681.6	46.0	7823	1 AB024741	AB024741 Comamonas
3	2490.6	42.7	9785	5 124403	124403 Sequence 1
4	2244.8	38.5	5029	1 PSETBMAF	L40033 Pseudomonas
5	1988	34.1	9511	2 AF026065	AF026065 Ralstonia
6	1949.2	33.4	8654	1 AB031996	AB031996 Ralstonia
7	1729.2	29.7	6112	1 PSEPEHEA	D28844 Pseudomonas
8	1708.8	29.3	6343	1 PPPHEHYD	X80765 P. putida ge
9	1617.2	27.7	5449	1 PSEPHHYD	M60276 Pseudomonas
10	1616.6	27.7	5700	1 PPPHH	X79063 P. putida ge
c	11 1213.4	20.8	2934	2 AF012632	AF012632 Ralstonia
	12 1043.4	17.9	1117	1 AB016861	AB016861 Burkholder
13	799.6	13.7	5552	1 D85083	D85083 Acinetobact
14	779.2	13.4	5905	1 ACPHENOL	Z36909 A. calcoacet
15	740.8	12.7	1154	1 AB017631	AB017631 Comamonas
16	723.8	12.4	1127	1 AB016863	AB016863 Comamonas
17	707.4	12.1	1133	1 AB016860	AB016860 Pseudomon
18	665.6	11.4	1144	1 AB016859	AB016859 Pseudomon
19	646	11.1	1143	1 AB016857	AB016857 Pseudomon
20	643	11.0	1139	1 AB016854	AB016854 Pseudomon
21	617	10.6	985	1 AB016858	AB016858 Pseudomon
22	455.4	7.8	3511	2 AF112137	AF112137 Burkholder
23	451.8	7.8	2640	2 CSU93090	U93090 Comamonas s
24	431.2	7.4	3103	2 AF109307	AF109307 Pseudomon

25 405.4 7.0 2191 1 PSETBUD
 26 393.2 6.7 614 1 AB016862
 27 376.6 6.5 615 1 AB016856
 28 376.6 6.0 646 1 AB016855
 29 334.2 5.7 684 2 AF228346
 30 330.6 5.7 1145 1 PPTDNC
 31 265.6 4.6 1575 2 BP020258
 32 261.2 4.5 1436 2 BC047111
 33 254.8 4.4 2433 2 AF112136
 34 236.2 4.1 1146 2 U01826
 35 201 3.4 6379 1 RERAMOA
 36 201 3.4 6379 5 E08269
 37 201 3.4 6379 3 I36963
 38 201 3.4 6379 23 E10607
 39 201 3.4 6379 23 E10980
 40 173.8 3.0 1575 2 BP020258
 41 152 2.6 8620 2 AF065891
 42 142 2.4 9037 1 PWMXYL
 43 141.8 2.4 209 1 AB011577
 44 140.2 2.4 209 1 AB011575
 45 138.6 2.4 209 1 AB011572

ALIGNMENTS

RESULT 1

AB006479 6298 bp DNA BCT 06-NOV-1998
 LOCUS Comamonas testosteroni gene for phenol hydroxylase component,
 DEFINITION ferredoxin like protein, catechol 2,3-dioxygenase, complete cds.
 ACCESSION AB006479
 VERSION AB006479.1 GI:3845544
 KEYWORDS aphB; aphQ; aphP; aphO; aphN; aphM; aphL; aphK; catechol
 2,3-dioxygenase; ferredoxin like protein; phenol hydroxylase
 component.
 SOURCE Comamonas testosteroni (strain:TA441) DNA.
 ORGANISM Comamonas testosteroni
 Bacteria; Proteobacteria; beta subdivision; Comamonadaceae;
 Comamonas.
 1 (bases 1 to 6298)
 Direct Submission
 Submitted (11-AUG-1997) to the DDBJ/EMBL/GenBank databases.
 Hirotsuki Arai, The RIKEN Institute, Laboratory of Microbiology;
 Hirotsuka 2-1, Wako, Saitama 351-0198, Japan
 (E-mail:harai@postman.riken.go.jp, Tel:048-467-9545,
 Fax:048-462-4672)
 2 (sites)
 Arai, H., Akahira, S., Ohishi, T., Maeda, M. and Kudo, T.
 Adaptation of Comamonas testosteroni TA441 to utilize phenol:
 organization and regulation of the genes involved in phenol
 degradation
 Microbiology 144 (Pt 10), 2895-2903 (1998)
 99018839 Location/Qualifiers
 source 1. 6298
 /organism="Comamonas testosteroni"
 /strain="TA441"
 /db_xref="taxon:285"
 283..474
 /gene="aphK"
 283..474
 /gene="aphK"
 /codon_start=1
 /transl_table=11
 /product="phenol hydroxylase component"
 /protein_id="BAA34169.1"
 /db_xref="GI:3845545"
 /translation="METSLLDLPVCDLTRRSRVLOKRVGFVDFEFSVGWPELTVEL
 AMTEADFEAFQCQVKARLL"
 516..1508
 /gene="aphL"
 /gene="aphL"

CDS
 gene
 CDS
 gene
 CDS
 gene
 CDS
 gene
 CDS
 gene

1508
 /gene="aphL"
 /codon_start=1
 /transl_table=11
 /product="phenol hydroxylase component"
 /protein_id="BAA34170.1"
 /db_xref="GI:3845546"
 /translation="MNIELSAREIQPLRHFTARVAAYTGDKPASRYLEALIGVQPTTH
 FHYPLMEPEFEDTGTARIMADNWLKDDPROYYTATMTTRAKOQEAWEANYQFV
 ESRNLAGKMSDALRAKACGVLMLRHHVANGNNMNSQICSRGYGTPTTAPAMHMAOH
 LGVAQYTRGLGVNDEPGVLEAGKNDWMQHPAQOPLRLHVEDLIVQDPIELFVAQNM
 VDLGLLPLVYNHFVDERVTVOGGTAVAMLTAFPEWHTETARWIDAVVKAASAD
 NRSQLLRWYLSYADHAGVALAAVAELGARGAVALSEVRQSLDIRAKGLDI"
 1520..1813
 /gene="aphM"
 1520..1813
 /gene="aphM"
 /codon_start=1
 /transl_table=11
 /product="phenol hydroxylase component"
 /protein_id="BAA34171.1"
 /db_xref="GI:3845547"
 /translation="METSASTPVSKVFIADFQNEESRPVVDAILADNTHAQAVYSPGL
 VKIDAPGRVLIRRASTIEQGRPYELQIQIINLVTLSGHVEDDEDDFSLSNQ"
 1861..3471
 /gene="aphN"
 1861..3471
 /gene="aphN"
 /codon_start=1
 /transl_table=11
 /product="phenol hydroxylase component"
 /protein_id="BAA34172.1"
 /db_xref="GI:3845548"
 /translation="MDAPVKKLGLKDRVTAMTRGLGWETSQPMKVPYPYDKFEG
 KIHDWDMEDPFRITMDAYWKYQGEKEKLYAVIEAFQANNQGLVTDARYLNALKLF
 IQGVPLEYYAHVGFAGHAGRHFTGAGARVAQMSIDELRHFTQETETHALSNYKFG
 MHSSHWFDVWFVSPKSPFDALSAPEFLTAVSFSEYVLTNLLFYPMFMSGAH
 NGDLSTVTFGSAQSDSRHMTLGICIKEMLEQDDPNVPIQVMDIKWFWRGYRVLT
 LVAMQDYMLPKRVMKWEAWMYAENGAGLFRDLARYGIREPAGHKHAEKGDHLS
 HQAWNIFYNTAAAPFTWVPGQEMQWLSSEKYPDSFRHRYRPLREHRAEQQAGKRF
 YSKTLPMLCTTCQIPMGFTPEGDATKICYRESYEGNKYHFCSDGCKHVEDDEPEKFV
 QSWLPVHQIYQGNCFKPDVPTAEGFDPLAAVLWDYGMNAGRONMDPEGSEDORNF
 WRGDPPTASANNAASSANSQEAQp"
 3468..3824
 /gene="aphO"
 3468..3824
 /gene="aphO"
 /codon_start=1
 /transl_table=11
 /product="phenol hydroxylase component"
 /protein_id="BAA34173.1"
 /db_xref="GI:3845549"
 /translation="MSVTALKPYRFFPAKDVRENPPAPLLYIGWEDHLMFCSPVALPLP
 ADMFPGALTQAVLPDYVYGAHPDFARIDMSEVQWFKSQPWPSPDASKSLADNGLTHKDV
 IREFTPGLIGIQGSYS"
 3884..4957
 /gene="aphP"
 3884..4957
 /gene="aphP"
 /codon_start=1
 /transl_table=11
 /product="phenol hydroxylase component"
 /protein_id="BAA34174.1"
 /db_xref="GI:3845550"
 /translation="MSYQLTLEPLGASIEVEEGOTLLDAALRQGIYIPHACGHLCGT
 CKIQVGDGDVHGGAANPFLMDMEREDGMTLACCATLQSDATIEADFPDQAIIPV
 RPDFAAVRIVETLPTTIKALHRLDOPMYQAGYVOLRIPGLSEOGGSRASIANA
 PGADGCAQETELNRLVLGCEGTGLWKHLSEGDSLQLAGPYGRFVHRSAROPMFM
 AGSGSLSSPRAMILELLANGCEQPIITLYQQRSEELYDDERALLAAQHPHTYIPA
 ISGDDCGDWGAGGFVHEAAQAFQGGFACGKATLCGPPPMVEACISTLMQGRUFEND
 IYTEKFLSDADAQARSPLFKRV"
 4965..5330
 /gene="aphQ"

[illegible]

QY	5143	ggcagctgcgaactgcaagtgcgcggcggtccaggaaagccgcttctt-----ctcggcca	5197
Db	5229	GAGGCCGTGCGCATTTGGCTGGCAGCGCCTGCACAAACCATTATTTTCTGCGGGCGCC	5288
QY	5198	tggcatgtgcgcggcacggcgatcaacaagttaaacaggagagagaacctaccattgggtgt	5257
Db	5289	TCCCCGGCGCCCAACAACCCAGCGCACACCAACAAGCAACTAAGGAGACAACATATGGGTGT	5348
QY	5258	gatcgatttggtcatgtcagttctgaagtgatggacatggaagcgcgctgcgtcatta	5317
Db	5349	AATGGAATCGCCACGCCCACTTCGCGCTGATGGACATGGATGCTGGCTCAAGCACTA	5408
QY	5318	cgtacggtgctcggcatgcaggaaacagatgcgcgacgcgcgggcgcaactctacctgaa	5377
Db	5409	CGAAAACCTGCTGGCATTAAGGTACAGTTCAGCTGCGCGATGTCAGCGGCAATGCTACCTCAA	5468
QY	5378	atgtgggacaaaatgggacaaagtattcgtgatcctgtcgcgttcgatacaggcgggct	5437
Db	5469	ATGTTGGACAGTGCGGACAAAGTAGTACTCCTCATCTCACGCAAAAGTATGTCGCGCGGCT	5528
QY	5438	caagcatcgccctacaaaggtcgagcagcgcgcatcttgtatgctgcgacgagcgcat	5497
Db	5529	CANTCATGTGGCTTACAGGTGCAAAACGATGCCGACCTCGACAGCTCGCAGGCACGCAT	5588
QY	5498	cgaagcttacgggatatcgacagagatgctcccgaagcgcgctgcgcggctcgccg	5557
Db	5589	CGAGSCCTGGGTTGTAAGACACCATGCTGCCGAAGGCACGCCACCGTGGGCGC	5648
QY	5558	caaactgcggttctgtgcgcagcggccatgaactgcggtgttcgcaagaagcgct	5617
Db	5649	CATGCTGCAGTTCGATCTGCCCTAGCGGCATGAGATGCGTGTACCCCAAGAAGSAGTG	5708
QY	5618	ggtgggcacgcggtlcggctcgctgaaccccgatccgtggccgacacatcccggtc	5677
Db	5709	CGTGGGCACGACGTGGGCTCTGTAACCCCAGCCCTTGGCCCGATGGTCTCAAGGCGC	5768
QY	5678	ggcgtgcactggctgcaccactgcctgctgatgtcgaaactgaacccggagcgcgct	5737
Db	5769	AGGTGCGCACTGGATAGACCACCTGCCTGCTGATGCGAGATGAATCCGAGGCGGTAT	5828
QY	5738	gaaccgctcgagagaaacagcgtctaatggcagatgtctgaacttcacttcgacga	5797
Db	5829	CAACACCGTGGAGGACAAACACCCGTTTCATGGCAGAGTGCATGGACTTCTCTCTGACCGA	5888
QY	5798	gcagtgatgtcgcccgcggaacacgat	5827
Db	5889	GCAGATTCGTGGCTGGCCCCCGCAAGCAATAT	5918
RESULT	2		
AB024741			
LOCUS	AB024741	7823 bp	DNA
DEFINITION	Comamonas testosteroni phcR, phcK, phcL, phcM, phcN, phcO, phcP and phcQ genes for regulator, phenol hydroxylase subunits and ferredoxin-like protein, complete cds.		
ACCESSION	AB024741		
VERSION	AB024741.1	GI:6505654	
KEYWORDS	ferredoxin-like protein; phenol hydroxylase subunit; positive regulator of phenol-degradative genes.		
SOURCE	Comamonas testosteroni (strain:R5)	DNA.	
ORGANISM	Comamonas testosteroni		
REFERENCE	Bacteria; Proteobacteria; beta subdivision; Comamonadaceae; Comamonas.		
AUTHORS	1 (sites)		
TITLE	Teramoto,M., Futamata,H., Harayama,S. and Watanabe,K. Characterization of a high-affinity phenol hydroxylase from Comamonas testosteroni R5 by gene cloning, and expression in Pseudomonas aeruginosa PAO1c Mol. Gen. Genet. 262 (3), 552-558 (1999) 2005/761 Direct Submission		
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			

JOURNAL

Submitted (10-MAR-1999) to the DDBJ/EMBL/GenBank databases. Maki Teramoto, Marine Biotechnology Institute Co., Ltd., Kamaishi Institute; Heita 3-75-1, Kamaishi, Iwate 026-0001, Japan (E-mail: mteramoto@kamaishi.mbio.co.jp, Tel: 81-193-26-6537, Fax: 81-193-26-6584)

FEATURES

source

Location/Qualifiers

source 1-7823

gene	CDS
gene	<p>1. 77025</p> <p>/organism="Comamonas testosteroni"</p> <p>/strain="R5"</p> <p>/db_xref="taxon:285"</p> <p>complement(297..2051)</p> <p>/gene="phcR"</p> <p>complement(297..2051)</p> <p>/gene="phcR"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="positive regulator of phenol-degradative genes"</p> <p>/protein_id="BAA87867.1"</p> <p>/db_xref="GI:6505655"</p> <p>/transl_table="MRVNPVARIETCFIDSFLMHPHTPALPSDADLRDSQVRFSTEDGLI</p> <p>WLSQRMILLHLASLALRLALRRENMNTMGPDTRILLRLRAGVAEGERADHLLRNPAS</p> <p>LFEMFAPQHLRLEGAVRATPEIFEAEQAGHFQCVRWHEHWEAEIHLLREWPQDO</p> <p>PSCMWLLGASGYSSAFRRVPEFKQSCGHAHGLIEGRFQHEMPDGEATARDY</p> <p>PDSMLRDLQSOVEALRTGLAPCDSQGLLGHSAFQAALLSKAAFTQVTVLIT</p> <p>GETGVGERARLHAMSPRADKPFVANCAALPADLIESELFGAEKAGFTGATATRI</p> <p>GRFRAGGRTLMDELCEPLAAQAKLLRVLOSSEVERLGSBHSFSTQVDVRIAAVND</p> <p>LEKAVEGRRRDLLYRLNYPPIRIPALRERVDDIELLAMLHLLQKXSALHGKAPGES</p> <p>DLAUTCAGDHPNGVRENLNVERGLITPAGEPIDVACLFQWSDNGQSTIDDOGH</p> <p>LQAACGADGDLTAGVYDVSHQHOGLSLDALEDGLLQAEVRRASGNLAAARALGMRTP</p> <p>QLSYLSRIERTTGTTEE"</p> <p>2306..2497</p> <p>/gene="phcK"</p> <p>2306..2497</p> <p>/gene="phcK"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="phenol hydroxylase subunit"</p> <p>/protein_id="BAA87868.1"</p> <p>/db_xref="GI:6505656"</p> <p>/transl_table="METSSTLDLPACDLTRSVRVLOKRANGFVDFEESVCGWPELSVEL</p> <p>AMTEADFEAFCCQVQKARLL"</p> <p>2539..3531</p> <p>/gene="phcL"</p> <p>2539..3531</p> <p>/gene="phcL"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="phenol hydroxylase subunit"</p> <p>/protein_id="BAA87869.1"</p> <p>/db_xref="GI:6505657"</p> <p>/transl_table="MNIELSARETOPLRTHFARVAAYTGDKPASRYLEATPLGVQPTH</p> <p>FHYRPLWEPPEFLDFTGRTAIRADNMALRDPROYYATWTMTRAKQOEAVEANYQV</p> <p>ESRNACKMSDALARACADVILMLRHHVAGWNNNSQICSRGYGTPTPTAAAMHAMDH</p> <p>LGVAQYITRLGLVMDPEGVLEACKNDWMOHPAQPLRHVEDLLVQDDPMLFLAONM</p> <p>VLOGLLPYLLNHFVDEBHVYVQSGTAVAMLTAFMPWHETETARWIDAVYKVAASEND</p> <p>NRQSLSRWLSYADHAGVAALAAVELALGVRGAAVSEVRQSLDARAKRAGLDI"</p> <p>3543..3836</p> <p>/gene="phcM"</p> <p>3543..3836</p> <p>/gene="phcM"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="phenol hydroxylase subunit"</p> <p>/protein_id="BAA87870.1"</p> <p>/db_xref="GI:6505658"</p> <p>/transl_table="MSTSASTPYSKVFIAPQDNEESPVVDAIADNTHAQAVYSPGL</p> <p>VKIDAPGRLVIRRASTIEEQTRPYELQIQILNVLTLGSHVDEDDDEFSLSWQ"</p> <p>3884..5506</p> <p>/gene="phcN"</p> <p>3884..5506</p> <p>/gene="phcN"</p> <p>/codon_start=1</p> <p>/transl_table=11</p>
gene	<p>3884..5506</p> <p>/gene="phcN"</p> <p>3884..5506</p> <p>/gene="phcN"</p> <p>/codon_start=1</p> <p>/transl_table=11</p>

/product="phenol hydroxylase subunit"	
/protein_id="BAA87871.1"	
/db_xref="GI:6505659"	
/translation="MDAPVIKKKLKDRYTAMTRGLGWETSYQPMDKVPFPYDKPFGI KIHWDKWDPPFLTMDAYWKYQGEKKLYAVIEAFQONNGOLGYTVDARYLNALKLF IQGVPLEYYAHGFAGHGHFTGCARVAQAQMSIDELRHFTQETHALSNKYKFG MHSSHWDFRWFLSLVPKSFSDFDALSGAFPEFLTAVTSFEIYVLTMLLFPVPMSSGAAR NGDLSYTFGSAQSDSEFFEDTIEICIKFMLEDSPGNVPIVORMIDKWFWRGVRVLT LWAMODYMLPKKVMSSKAWEMAEENGALFDRLARYGIREPACGWLACBCKDHS HOAVNIYFNATPHTWVPGQEMWLSEKYPDSFDKHYRPLRLEHYLAEOQAGRFV YSTLMLCTTCOIPMGTEPGDATKICYRESDYEGSKYHFCSDGCKHYVDEDEPKRY QSWLPHQIYQGNCFKPDVDPTAEGFDPLAASVLDWYGMVGRDNMDFEGESDQRNPAA WHGDDPPATYSANNAASNAASADNPSEQAQS"	
gene	5503.5859
CDS	/gene="phcO" 5503.5859 /gene="phcO" /codon_start=1 /transl_table=11 /product="phenol hydroxylase subunit" /protein_id="BAA87872.1" /db_xref="GI:6505660" /translation="MSVTALKPYRFPKDVRENFPAPLLYIGWEDHLMFCSPVALPLP PDMFGALQAQVLPDYVGAHPDEARIDNSVQWFKSGQSWSPASKSLADNGLITHKDV IRPFTPLGIGIQGSYS"
gene	5919.6992
CDS	/gene="phcP" 5919.6992 /gene="phcP" /codon_start=1 /transl_table=11 /product="phenol hydroxylase subunit" /protein_id="BAA87873.1" /db_xref="GI:6505661" /translation="MSYQLTLEPLGASIEVEGOTLLDAALRGQIVIPHACGHGLCGT CKYQCGDGEVDHGAANPFAIMDMERDGMTLACCATLQSDATIEADFEPAQIIPV RDFAAVTVRIYDUTPIKALHLRDSMRYQAQYVOLRIPGLSEQGSRAFSANA PGADGCAETELNVLPGAGYGMUHLHSEGLQLAGYGRFVFRHSAACOPWVF AGGSGISPRAMILLANGCEQPIITLVYQGRSRELYDDDEFRAALAAOHPHFTYIPA ISGEDCGWQAGRFVHEAAQAFQGGFAGCKAYLCGPPPMVEACTISTLMQGRLEFRD IYVEKFLSAADAGQARSPLFKRV"
gene	7009.7365
CDS	/gene="phcQ" 7009.7365 /gene="phcQ" /codon_start=1 /transl_table=11 /product="ferredoxin-like protein" /protein_id="BAA87874.1" /db_xref="GI:6505662" /translation="MONVQVSVBEQTDYACGTHESLLSMLRLGRKKGIPVCGVNGGC GVCKVQVLEGAVRHLGPVSCAHVSDLERDQGYTLACRVAPLEAVRIAQRLLHKPFPL RAASPAPTTHDPTSN"
BASE COUNT	1479 a 2488 c 2448 g 1408 t
ORIGIN	
Query Match 46.0%; Score 2681.6; DB 1; Length 7823; Best Local Similarity 69.5%; Pred. No. 2.2e+265; Matches 3850; Conservative 0; Mismatches 1604; Indels 86; Gaps 12;	
Qy	228 ccgatcatgaatcagaccaccacccgatctttcccgcttcgatcccgccgaagtgcgc 287
Db	2300 CCCCTCATTGGAACCAACGACTCTCGATCTTCGCCGCTGCGCATCTACGCCGCCGTCGGTG 2359
Qy	288 ccgctgaccgcagcaacgcgcgcgcgcttcgctgaattcgagctgcatcgccgcgcgc 347
Db	2360 CGCGTGTGCAAAAGCGGCCACAGCGTTTGTGGACTTCGAGTCTCCCGTGGCGTG--G 2416
Qy	348 ccggaactgtgcgcgagctgaogttgtctctgcgcgcttcgatcgcttcgcgcgaa 407
Db	2417 CCCGAACCTCTCGTGGAGCTGGCCATCACAGAGCCGACATTTGAGGCGCTTCTGCCAGGTC 2476

	Query Match	46.08;	Score 2681.6;	DB 1;	Length 7823;
	Best Local Similarity	69.3%;	Pred. No. 2.2e-265;		
	Matches 3850;	Conservative	0;	Mismatches 1604;	Indels 86; Gaps 12;
Qy	228	ccgatcatgaatcagcagccacccagcatcttcccgcttcgatccggcgccagatgcgc	287		
Db	2300	CCCTCATGGAACACAGACTCTCGATCTTCCGCCCTGGCATCTCAGCGCCGCTGGTG	2359		
Qy	288	cgcgtgacggcgacgaacgcgcgcgcgttcctgcgaattcagctgttcgatcgcgcgcgcg	347		
Db	2360	CGCGTGTGTCNAAGCGGCCACAGGTTTTGTGACTTCGACTTCGCCGGGGCTG--G	2416		
Qy	348	ccggaactgtgcgcgagctgaagttgtctctcgcgcgcttcgatcgcgttcctgcgcgaa	407		
Db	2417	CCCCAACTCTCGGTGGAGCTGGCCATCACAGAGCCGCACTTTCGAGGCTTCTGCCAGGTC	2476		

[illegible]

3553	Db		CTGTAGCACCCCGTCTCCAAGGTGTTCATCGCCTTTTCAGGACAAACGAGGAATCGCGCC	3612
1541	Qy		cgatcgtggatgcatcgctcgcgcaaacccgcgcgcgtgtgtcagtcagtcgccgca	1600
3613	Db		CCGTGGTGGAGCCCATCTCGCGCAACACCCACGCCCGGCGGTGTACTCGCCTGGTC	3672
1601	Qy		tgttcaagatcgacgcgcgcgcgcgtgacgatccgcgcgcaaacgatctgaggaactga	1660
3673	Db		TCGTGAAGATCGAGCGCCCGCGCGCTGTGTATACGTGCGCCAGCATTTGAAGAGCAGA	3732
1661	Qy		ccggcacgcgcttcgaacctgacagctcccaagtcgaacctgatacgcgtgtcaaggcca	1720
3733	Db		CCGTTCGCCCCATAGCAGCTCGACGATCCAGATCAATCTGTGTCACTGCTCGCGCCATG	3792
1721	Qy		tcgacgagcagcagcagagttcacgctgaagcttgctcactgaacgcgcgcgcgcacgcgc	1780
3793	Db		TGGATGAGGACGACACGAATTTCTCCTGAGCTGGAAACCATGTAAGAGCGCTTCACCGAAC	3852
1781	Qy		accgacaacacccggagacaga-----atggacacgcgaacgcgtcaagaaaaaact	1831
3853	Db		ACCACGACAAACCCATCCAAGGAGACACCATGGAGCACCGTAAATCAAGAAAAAGCT	3912
1832	Qy		cggctgaagacacgcgtacgcgcgaatgacgcgcgcgtcgctcgtggagacacactaca	1891
3913	Db		GGGCTCAAGGACCGCTACACCGCCATGACACGTGGCCTTGGGTGGGAGACGACTACA	3972
1892	Qy		gccgatggacaaggtcttccgttacgacgcgtacgagggcatcaagatccacgactggga	1951
3973	Db		GCCCATGGATGAAGTCTTCCCTATGACAAGTTCGAGGGGCATCAAGATCCAGACTGGGA	4032
1952	Qy		caatgggtgacccgttcgcgcgtgacgatggatcgctactggaaataccagggcgagaa	2011
4033	Db		CAAGTGGGAAGACCCCTTCGCTGACCATGGAGCGCTACTTGGAAGTACACAGGGCGAAA	4092
2012	Qy		ggaaaagaagctgtacgcggtgatgacgcggttcacgcgaacaacgcgcttcctcggcgt	2071
4093	Db		AGAGAAAAAGCTCTACGCCGTGATCGAGGCCTTTGCCAGAAACACGGCCAGTTGGGCGT	4152
2072	Qy		gagcgacgccctacatacaacgcgtgaagctgttcctccagggcgtgacgcgcgtcga	2131
4153	Db		GACCGATGGCGGTATCTCAACGGGCTCAAGCTGTCATTAGAGGCGGTGGTGGCGCTGGA	4212
2132	Qy		atactgcgcacgcgcggttcgcgcgtatgcgcgcgcgacttcacgcgcgagggcgcgcg	2191
4213	Db		GTATTAGCCCAACCGCGGTTTTGCCATCGGCGCGGCATTTACCGGTGCGGGGCGACG	4272
2192	Qy		catcgcgtgcagatcgatcgacgcgctgcgcgaactacgcgaacgcgaacgcgactgc	2251
4273	Db		CGTGGCGCGCGATGACGATGCCATCGACGAGCTTGCGCCACTTTCAGACCGAGACCCATGC	4332
2252	Qy		gatctgacgtacacaagttcttcacgggttcatacactcgaaccagtggttcgacgcg	2311
4333	Db		GCTGTCCAACTACAAACAGTACTTCAACGGCATTCACACGCTCCAGCCATGTGTCGACCG	4392
2312	Qy		cgttggtaccctgcggtgcgcgaagtcgttcttcgagagcgcatctcgtcggggcgctt	2371
4393	Db		GGTCTGGTTTTCTGCGTGGCCCAAGTCTTTTGAAGATGCGCTGAGCGCAGGCGCCGTT	4452
2372	Qy		cgagttcctgacgcgcggttcgagcttcgttcgaatacgtgtcgaacacctgcgttgcgt	2431
4453	Db		CGAGTTCTCAGGCGAGTGAGCTTCCTTCGAATATGTGTGACCAATCTGCTGCTTCGT	4512
2432	Qy		gcggttcagtgcgggcgcgcgtctacaacgcggtgacatgtgcacgcgtcaagcttcctc	2491
4513	Db		GCCCTCATGTGCGGGCGGCGACAAACGGGATCTGTCCACCGTGTTCGCTTGGCTTCTC	4572
2492	Qy		cgcgcagtgcgagcaatcgcgttcacatgacgtcgcgcacatcgaatcatcaagttccctgc	2551
4573	Db		GGCCCAAGCGATGAGTCTCGCCCATATGACGCTGGGCATTCGAATGCATCAAGTTTCATGCT	4632
2552	Qy		cgaacaggacccggacaacgcgcatcgctgcgcgcgtgcagcgcgtgga tcgacaagtggttcggcg	2611

[illegible][illegible]

Db 5018 AACGGGTTTGA 5029

RESULT 5
AF026065

LOCUS
DEFINITION Ralstonia sp. E2 positive phenol-degradative gene regulator (poxB),
phenol hydroxylase components (poxA, poxB, poxC, poxD, poxE, poxF),
and ferredoxin-like protein (poxG) genes, complete cds.

ACCESSION AF026065
VERSION AF026065.1 GI:3445530

KEYWORDS
SOURCE Ralstonia sp. E2.
ORGANISM Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
Ralstonia.

REFERENCE 1 (bases 1 to 9511)
AUTHORS Hino,S., Watanabe,K. and Takahashi,N.
TITLE Phenol hydroxylase cloned from Ralstonia eutropha strain E2
exhibits novel kinetic properties
JOURNAL Microbiology 144 (Pt 7), 1765-1772 (1998)

MEDLINE 98361023

REFERENCE 2 (bases 1 to 9511)
AUTHORS Watanabe,K. and Hino,S.
TITLE Direct Submission
JOURNAL Submitted (18-SEP-1997) Kanaishi Laboratory, Marine Biotechnology
Institute, Heita 3-75-1, Kamaishi, Iwate 026, Japan

FEATURES
source location/Qualifiers
1..9511
/organism="Ralstonia sp. E2"
/strain="E2"
/db_xref="taxon:67939"
2014..3720
/gene="poxB"
2014..3720
/gene="poxB"
/note="PoxB"
/codon_start=1
/transl_table=11
/product="positive phenol-degradative gene regulator"
/protein_id="AAC32451.1"
/db_xref="GI:3445531"

translation="MSSSTDNFSAWTRDGLNLARRLRFAMKEGSIWLGQRMILLHT
AALGRLKELMTGERARGLRGMFGHSGVRDAELAATMRSGHSDFGMLEMGPCILH
TTEGVVTPVTIAAGVYHGFEIWEDESGDVHRQMFVAQAPVCWMQIGYATG
VTSALMGKITLYRELECVCGCHPHCRILGKPLOEWDEGEFLAYOPDPVIDITLAIQ
SEVEOLRALORDPADLVGSGPCFRFAWNLLORAAGSVTVLLGCTGCKERFAQ
ALHUGSARAGFPFAVNCAAPDELIESELFGKGAFTGHQSRAGRFRAGGGTLF
LDLIELSASAKLLVLQGEVERGVGNEARVDRLVAATNVDLAEAVROGTFRK
DIYLRNVPVPTPLRELDDIRLAERFYARGHKIKILGITDRALAEALLRYDM
PNVRELENIBERGVIILAANGSOISAEQLFLPGAEPAPVDVTTPLRGEOQALPLREA
AVHGLDYMAENGLALGESVESYLMEAAQORADGNLSQAARLLGLTRLQLAYRWKTRGT
RG"

gene 4490..4714
CDS /gene="poxA"
4490..4714
/gene="poxA"
/note="PoxA"
/codon_start=1
/transl_table=11
/product="phenol hydroxylase component"
/protein_id="AAC32452.1"
/db_xref="GI:3445532"

translation="MINGSVPLFEATPRYVRVEGRTPEGFVLFPAFSVADPDNLVELLI
PRPMFEAFCSANHVRFLPAQACADSGEDA"

gene 4807..5802
CDS /gene="poxB"
4807..5802
/gene="poxB"
/note="PoxB"
/codon_start=1
/transl_table=11
/product="phenol hydroxylase component"

/protein_id="AAC32453.1"
/db_xref="GI:3445533"

translation="MOVDIKTOOIQPIROTGYGHVARFEGDKPASRYOEATYDVQSEVN
PHYRTWAQPELYDKRRTALEMADWYALKPRQYYGVAGVSTGRQOEAEEKNEFAFV
EKRGLOALPLEWRERLIDGLPLRHVENAANNFNFCADYGWGCTALTQACTYCAMD
LGIAOYLSRIIGLLDNGTAALERARAWEASEAWQPLRRFEHSFVIEDWTFOTVYTQ
NLALDGLYPYIVYHADAAIVRACGTGLAVVFTEFMDNRDHEVRWWDVAVVATAAESE
ANRALLSWAGEARAQAAREALRPVAAIILLGEGEQAITALCLEQFDARLARLGLVAA"

gene 5813..6085
CDS /gene="poxB"
5813..6085
/gene="poxB"
/note="PoxB"
/codon_start=1
/transl_table=11
/product="phenol hydroxylase component"
/protein_id="AAC32454.1"
/db_xref="GI:3445534"

translation="MSANYIALQNNDTRPIETAETAEANPLAVVSQFPAMVKIDAPG
RLTIIVRELVAADKLGRDWDLQEIHLNLSISLGNIDETDEAFTLHWSA"

gene 6153..7667
CDS /gene="poxB"
6153..7667
/gene="poxB"
/note="PoxB"
/codon_start=1
/transl_table=11
/product="phenol hydroxylase component"
/protein_id="AAC32455.1"
/db_xref="GI:3445535"

translation="MDARKKLNLREKYATMTRDLGWETTYPMDKVFPFKYEGIKIH
DWDKWDPMNTDAYWKYQSEKERKIYAIDSFVQNNHNLNSDARYLNALRLFLTG
VTPLEYAAHRYGAHLGRHFAGARVAAMOQSIDELSHAQTQLHTLSVKNKYFHGFGE
WRMHDRVMYLSVPKSYFEDAMSAGPEFTATSDFEYVLTNLLPFMSGAAYNGD
NATYTFGESAQSDSRHMTLGLLEVVKFLCEDDPNIPLLQKLDKFWRGFRLLLTILG
MMDMPLPRVMSNAEWEMTFEQAGALFKDLERYGLRYPKTYHDVATKTOKRTHSEA
WGTIFYAAAAAGFHTWPKPDEMAMWLDEKYPTQFYARYKPLRDYQOQOAGERFYNG
TLPMLCOTQIPMVFSPPDPTQTCYRESSYHGMKFHCSDGCKDIFDGPEKYAQAAQ
LPVHQIYOGNCGGASLEDVLRNLNLAGDNLDPFEGSQDKNNANKWGVHPHAA"

gene 7004..8063
CDS /gene="poxB"
7004..8063
/gene="poxB"
/note="PoxB"
/codon_start=1
/transl_table=11
/product="phenol hydroxylase component"
/protein_id="AAC32456.1"
/db_xref="GI:3445536"

translation="MSVASIGEYFEFPADREAVFHGNRLLYVGWDQHLLFCAPHCFPF
PPAMMRVWVQVLPVGVYHPDGRIDWSRVWLRCGRWPQDLDATLEENGSLGKHE
VIRPFRGLDGIIGGSAS"

gene 8089..9159
CDS /gene="poxB"
8089..9159
/gene="poxB"
/note="PoxB"
/codon_start=1
/transl_table=11
/product="phenol hydroxylase component"
/protein_id="AAC32457.1"
/db_xref="GI:3445537"

translation="MYSLTTEPIGOTIPTIAPGTQVTDACLURSGVWLPACCHCLCATC
KVQVVEGDQGEASSPALMDFEDNGOCACCATQASOLDTIEADIIEADALGLPLA
DYAEVDVRAITPTTIRGILRVRKGAKVAFOAGYLNQLQVPCGDOPRAFSLANAPDE
ELVELHVRRKVEGGATGYLHERLAVGDALRFSAPYGRFFVRSQAQMPYLPALSDEP
SPRAMIRELLAAGETLPTLVQGARNRAELYGEDEFRALAERHPNRYYPALSDPEAD
SAWQARGFAHAALAAALYADGDGRADRFRGHKAYLCGPGPPPMEIACIRTLMQGRLEADI
HTEKFLSADSAQNSARSPLEFI"

gene 9205..9510
CDS /gene="poxB"
9205..9510
/gene="poxB"

```
/note="PoxG"
/codon_start=1
/transl_table=11
/product="ferredoxin-like protein"
/protein_id="AAC32458.1"
/db_xref="GI:344538"
/translation="MHTVETADSGQRYPCDPGPNLLRAMEVLGQRIIPAGCRGGGCV
CKVRIESGRYRTKMSRACLSAEQGLVLAKAFPSDILRLPAALLARCLDKAR"

BASE COUNT   1592 a   3268 c   3265 g   1386 t
ORIGIN

Query Match          34.1%   Score 1988;   DB 2;   Length 9511;
Best Local Similarity 66.0%;   Pred. No. 9,4e-195;
Matches 3202;   Conservative 0;   Mismatches 1470;   Indels 182;   Gaps 15;

QY  394  cgttctcggaacagcaggttcacgcggttcgacgtcgaagcgaaccccatgaccttgag 453
DB  4738  CGGCGCCGACAGCCCTCACCGCGCGGACGCGCGCCGCCACACGATACAAACACGGA 4797

QY  454  gagcaagaagtgaacctcagcgtgaagaagatgacatacgaacgcgctccggcacacattt 513
DB  4798  GACAACCGATGCAAGTCGATATCAAGACGACGACAGATCCAGCGGCTGCGCCACGACATC 4857

QY  514  gcgcattgcgcgagaacatcgcgcgcgaagaacgacgcgcgcgtaccaggaaggcatg 573
DB  4858  GGCCACGTGGCGGACGCTT---CGGCGACAAAGCCGCCCTCGCGCTACGAGGAGGCGACC 4914

QY  574  atggggcgcgaagcccccaggagaacttcattacgcggccggaacctgggacccggactacgag 633
DB  4915  TACGACGTCCAGTCCGAGGTCAACTTCACCTACGCGCCGACCTGGGCGCGCGAGTTCGAA 4974

QY  634  atcttcgacgtcgcgtcgcgcgacatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 693
DB  4975  CTGTACGACAAAGCCCGCCACCGCCCTCAGATGCGGACCTGGTACGCGCTGAAGGATCCG 5034

QY  694  cgccagttctactacgcgtcgtggcgaccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 753
DB  5035  CGCCAGTACTACTAGCTGCTAGCTCGGTACGCGCGCGCGCGCCACGACGAGGAGCGCCGAG 5094

QY  754  tcgaacttcagttcgtcgaatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 813
DB  5095  AAGAATCTCGCTTCTGTCGAGAGCGCGGCGTCTGTGAAGCGTGTGCGCTGGAATGCGC 5154

QY  814  qcgcgggcgctcagcgtcgtcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 873
DB  5155  GAGCGCTGACCGATGCGCTGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 5214

QY  874  aacgcgcagatctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 933
DB  5215  AACTTCTACTTGGCGGACTATGCTTGGGCGACGCGCCATACGACGAGGCTGCACCTACTGCG 5274

QY  934  gcgatgaacacctcgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 991
DB  5275  GCGATGACGCGGTGGCGATCGCCCAATACCTGTGCGGCATCGGCGCTGCTGCTGCGACGCG 5334

QY  992  ----agcccaactgctcgaagggcccaagcgaactcgaacccgcgcgcgcgcgcgcgcgcgcgcgc 1047
DB  5335  AACACGGGCGTGGCTGGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5394

QY  1048  ccgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1107
DB  5395  CCGCTGCGCGCTTCTGTCGAGCAGACAGCTTCGTATCGAGGACTGGTTCCAGACCTTCGTT 5454

QY  1108  gcgcgaacacctcgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1167
DB  5455  ACGCAGAACCTGCGCGCTGCGACGGCGCTCTATCCGCTGGCTTACCAAGC---ATGCGGAC 5511

QY  1168  gaacgcgacgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1227
DB  5512  CGGCGCATGTCGCGCGCTGCGGCGACCGCGCTGGCGCGTGGTACGGAATTCATCAAGACGAC 5571

QY  1228  tggcacaccagtgtaaacccgtggtgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1287
```

```
DB  5572  TGGCGCGACGAGCAGCTGCGCTGGGTGATGCGGTGTGCAGAGCGCGCGCGAGTCC 5631

QY  1288  gagcaaacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1347
DB  5632  GAGGCCAACCGCGCGCTGCTGCTGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 5691

QY  1348  qcactggcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1407
DB  5692  GCGCTGCGCGCGGTGGCGGCCATCCTGCTGGGTGGCGAAGAGAGAGAGAGAGAGAGAGAGAGAG 5751

QY  1408  gtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1467
DB  5752  TGCCTCGAACAAATTCGACCGCGCGCTGCGCCAAAGCTGCGCGCTGGCGCGCTGANGA----- 5805

QY  1468  tctcccttaaccaagaatgccagcatgccaaacgtattcatcgcctttcagggccaat 1527
DB  5806  -----GACCGCCATGAGCGCCAAACGCTACATCGCCCTGCAGAACAAAC 5848

QY  1528  gaggaactccagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1587
DB  5849  GAGGACACCCGTCCTCATATCGAGGCCATCGCGAGGCCAACCGCTCGCGCGTGGTGTGCG 5908

QY  1588  gactgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1647
DB  5909  CAGTTTCCCGCCATGCTCAAGATCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGCTG 5968

QY  1648  atcgaggaactgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1707
DB  5969  GTGGCCGACAAAGTTCGCGCGCGCTGCGGACCTGCGAGAGATCCACCTGAACCTGATTCG 6028

QY  1708  ctgcagggccacatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1767
DB  6029  CTGTCCGGCAACATCGACGAGACGCGCAAGCCCTTACGCTGACCTGACGCGCGCTGAGCA 6088

QY  1768  ccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1823
DB  6089  CGGCTCGCGCTCGCGACACAAAGCAACGCGCAACGCGACGCGCATACCGAGAGCATACGGAGA 6148

QY  1824  -----aaaaactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1867
DB  6149  CATCATGACGACGCGAAGAGCTCAACCTGCGCGGAGAAAGTACGCGACGATGACGCGCGCA 6208

QY  1868  cctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1927
DB  6209  CTTGCGGTGGGAAACCACTTACGAGCGCATGGACAAAGTCTTTCCCTTCGACAAAGTACGA 6268

QY  1928  gggcacaagatcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1987
DB  6269  AGGCATCAAGATCCACGACTGGGACAAAGTGGAGGATCCCTTCGCGCATGACCATGGAGCG 6328

QY  1988  gtactgaaatacacggcgagagaaagaaagctgtacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2047
DB  6329  GTACTGGAAGTACAGTCGCGAGAGAGAGCGCAAGCTAGCTAGCGCATCATCGACTCTCTCGT 6388

QY  2048  gcagaaacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2107
DB  6389  GCAGAACACGCGCACCTCAACGCTTCGAGCGCGCGCTACTGNACCGCGCTGCGCGCTGTT 6448

QY  2108  cctccagggcggtgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2167
DB  6449  CCTGACCGCGGTGACGCGCTGGAGTACGCGCGCGACCGCGGCTATGCCACCTGGGGCG 6508

QY  2168  gaacttcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2227
DB  6509  CCATTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6568

QY  2228  gcactacagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2287
DB  6569  CCACGCGCACGCCAGTTGACACGCTGTCGGTCTACAAACAGTATTTCCACGCGCTTCGG 6628

QY  2288  tcactogaacacgtggttcgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2347
```


RESULT	7	
LOCUS	PSEPHEAA	6112 bp DNA BCT
DEFINITION	Pseudomonas putida phe[A1, A2, A3, A4, A5, A6] genes for phenol hydroxylase components, complete cds.	07-FEB-1999
ACCESSION	D28864	1 GI:840700
VERSION	D28864.1	GI:840700
KEYWORDS	phenol hydroxylase.	
SOURCE	Pseudomonas putida (isolate:BH)	DNA.
ORGANISM	Pseudomonas putida	
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group; Pseudomonas.	
AUTHORS	1 (bases 1 to 6112)	
TITLE	Takeo, M.	
JOURNAL	Direct Submission	
REFERENCE	Submitted (07-MAR-1994) to the DDBJ/EMBL/GenBank databases.	
AUTHORS	Masahiro Takeo, Himeji Institute of Technology; 2167 Shosha, Himeji, Hyogo 671-22, Japan (E-mail: takeo@chem.eng.himeji-tech.ac.jp, Tel: 0792-67-4893, Fax: 0792-67-4891)	
TITLE	2 (sites)	
AUTHORS	Takeo, M.	
JOURNAL	Unpublished (1995)	
REFERENCE	3 (bases 1 to 6112)	
AUTHORS	Takeo, M., Maeda, Y., Okada, H., Miyama, K., Mori, K., Ike, M. and Fujita, M.	
TITLE	Molecular cloning and sequencing of the phenol hydroxylase gene from Pseudomonas putida BH	
JOURNAL	J. Ferment. Bioeng. 79, 485-488 (1995)	
COMMENT	On Jun 1, 1995 this sequence version replaced gi:468465. Sequence updated (24-May-1995) by: Masahiro Takeo.	
FEATURES	Location/Qualifiers	
Source	1..6112	/organism="Pseudomonas putida"
gene	/isolate="BH"	/db_xref="taxon:303"
CDS	746..1024	/gene="pheA1"
	746..1024	/gene="pheA1"
	/codon_start=1	/transl_table=11
	/product="one component of phenol hydroxylase"	/protein_id="BAA06014.1"
gene	/db_xref="GI:468466"	/translation="MTVTNPTPTDOLTRYIRVRSEPEAKVFVEFDAFCHPELFLVEL
CDS	1077..2072	/gene="pheA2"
	1077..2072	/gene="pheA2"
	/codon_start=1	/transl_table=11
	/product="one component of phenol hydroxylase"	/protein_id="BAA06015.1"
gene	/db_xref="GI:468467"	/translation="MSVEIKTNTVDPIRQYTNLQRRFGDKPASRYQEAASYDIEAVTN
CDS	2076..2348	/gene="pheA3"
	2076..2348	/gene="pheA3"
	/codon_start=1	/transl_table=11
	/product="one component of phenol hydroxylase"	/protein_id="BAA06016.1"
gene	/db_xref="GI:468468"	/translation="MTVTNPTPTDOLTRYIRVRSEPEAKVFVEFDAFCHPELFLVEL
CDS	2348..2612	/gene="pheA4"
	2348..2612	/gene="pheA4"
	/codon_start=1	/transl_table=11
	/product="one component of phenol hydroxylase"	/protein_id="BAA06017.1"
gene	/db_xref="GI:468469"	/translation="MTVTNPTPTDOLTRYIRVRSEPEAKVFVEFDAFCHPELFLVEL
CDS	2612..2886	/gene="pheA5"
	2612..2886	/gene="pheA5"
	/codon_start=1	/transl_table=11
	/product="one component of phenol hydroxylase"	/protein_id="BAA06018.1"
gene	/db_xref="GI:468470"	/translation="MTVTNPTPTDOLTRYIRVRSEPEAKVFVEFDAFCHPELFLVEL
CDS	2886..3150	/gene="pheA6"
	2886..3150	/gene="pheA6"
	/codon_start=1	/transl_table=11
	/product="one component of phenol hydroxylase"	/protein_id="BAA06019.1"
gene	/db_xref="GI:468471"	/translation="MTVTNPTPTDOLTRYIRVRSEPEAKVFVEFDAFCHPELFLVEL
CDS	3150..3414	/gene="pheA7"
	3150..3414	/gene="pheA7"
	/codon_start=1	/transl_table=11
	/product="one component of phenol hydroxylase"	/protein_id="BAA06020.1"
gene	/db_xref="GI:468472"	/translation="MTVTNPTPTDOLTRYIRVRSEPEAKVFVEFDAFCHPELFLVEL
CDS	3414..3678	/gene="pheA8"
	3414..3678	/gene="pheA8"
	/codon_start=1	/transl_table=11
	/product="one component of phenol hydroxylase"	/protein_id="BAA06021.1"
gene	/db_xref="GI:468473"	/translation="MTVTNPTPTDOLTRYIRVRSEPEAKVFVEFDAFCHPELFLVEL
CDS	3678..3942	/gene="pheA9"
	3678..3942	/gene="pheA9"
	/codon_start=1	/transl_table=11
	/product="one component of phenol hydroxylase"	/protein_id="BAA06022.1"
gene	/db_xref="GI:468474"	/translation="MTVTNPTPTDOLTRYIRVRSEPEAKVFVEFDAFCHPELFLVEL
CDS	3942..4206	/gene="pheA10"
	3942..4206	/gene="pheA10"
	/codon_start=1	/transl_table=11
	/product="one component of phenol hydroxylase"	/protein_id="BAA06023.1"
gene	/db_xref="GI:468475"	/translation="MTVTNPTPTDOLTRYIRVRSEPEAKVFVEFDAFCHPELFLVEL
CDS	4206..4470	/gene="pheA11"
	4206..4470	/gene="pheA11"
	/codon_start=1	/transl_table=11
	/product="one component of phenol hydroxylase"	/protein_id="BAA06024.1"
gene	/db_xref="GI:468476"	/translation="MTVTNPTPTDOLTRYIRVRSEPEAKVFVEFDAFCHPELFLVEL
CDS	4470..4734	/gene="pheA12"
	4470..4734	/gene="pheA12"
	/codon_start=1	/transl_table=11
	/product="one component of phenol hydroxylase"	/protein_id="BAA06025.1"
gene	/db_xref="GI:468477"	/translation="MTVTNPTPTDOLTRYIRVRSEPEAKVFVEFDAFCHPELFLVEL
CDS	4734..5000	/gene="pheA13"
	4734..5000	/gene="pheA13"
	/codon_start=1	/transl_table=11
	/product="one component of phenol hydroxylase"	/protein_id="BAA06026.1"
gene	/db_xref="GI:468478"	/translation="MTVTNPTPTDOLTRYIRVRSEPEAKVFVEFDAFCHPELFLVEL
CDS	5000..5264	/gene="pheA14"
	5000..5264	/gene="pheA14"
	/codon_start=1	/transl_table=11
	/product="one component of phenol hydroxylase"	/protein_id="BAA06027.1"
gene	/db_xref="GI:468479"	/translation="MTVTNPTPTDOLTRYIRVRSEPEAKVFVEFDAFCHPELFLVEL
CDS	5264..5528	/gene="pheA15"
	5264..5528	/gene="pheA15"
	/codon_start=1	/transl_table=11
	/product="one component of phenol hydroxylase"	/protein_id="BAA06028.1"
gene	/db_xref="GI:468480"	/translation="MTVTNPTPTDOLTRYIRVRSEPEAKVFVEFDAFCHPELFLVEL
CDS	5528..5792	/gene="pheA16"
	5528..5792	/gene="

Db 3734 CGCCGAGCACCAGCGCTGGCTGGCCCTGAAAGGTTCAGACCCCAACCACTGCGCGCCCGG 3793
QY 3334 gccaggccagcgcgaactgacccgcaacgacaagaacaatttgaacgagggccgcgaagcg 3393
Db 3794 CGGACAGAGCCCTGGACGCGCCCTGAGGACGCGCCGCTCAGGGGTGAAGACACCGCC 3853
QY 3394 cggatcggaacgcggcgccgacagagagaaacatggcgtcatcgctcaaaaccta 3453
Db 3854 CATTCCAAAGAACAGAGGGTTCGATCATGACTGCTCAACTCAATCGGCAATA 3912
QY 3454 cgaactccgggtgaagatgcggtcgagaagtcttcgg---cgccgctgctcttaogtgg 3510
Db 3913 CCCCGCACGCCACGGGATGTGACGGCAACTTCAACGGCATGCAACTGCTTACCTCTA 3972
QY 3511 ctggaaacaccatctatgttcccgccgctctgtctgcgctgcgcccgaatgcc 3570
Db 3973 CTGGGAAGAGCACTGATGACTGCTCCGCGCTTCGGTGTCTTGGTAGCCCCCGGATGCC 4032
QY 3571 gttcggcgctggcgccgagcgtgctgcgcgcgtctacggctatacccccgaacttcgc 3630
Db 4033 CTTTGGCGAGTTCCTCGACGAGGTGCTCAAGCCCGGATCCACGCCCATCCGACAGCGC 4092
QY 3631 gaagatcgcactgggagtcgctgcgagtggttcggttcggcgagccggtggcccggaacc 3690
Db 4093 GAGGATCGATTTTCAGCCAGCGCGCTCTGGCAGGTGAACGACCAAGCGCTTCAACCCGACTA 4152
QY 3691 ggcgaagacctggccgcaacgacctcgggacaaaggaacctgatcagcttcgcgaagcc 3750
Db 4153 CGCCGCGACCTGGGAAGCAAGCGCATGACGACCAAAAGCATGCTGCGTCTGAACACCCC 4212
QY 3751 cggcctcgagcgcctcgccgcgcgagcttctgaccgcacgcgagcagcgaaacataca 3810
Db 4213 GGGCCTGAACGGCATCCAGGTTCTGTGACGCTGAGAGGTGTG-----TCA 4257
QY 3811 tgagccaccaacttaccatcgagccgctggcgctcagcagatcgaggtcgaggaagacaga 3870
Db 4258 TGACTTTACAACTGACATCGAAGCTTACCGGTGAATCATCATGAGGTGCGAGGAGGCGCAGA 4317
QY 3871 cgaatcgtatccgctgcgcagggacatctacattccgacgcgctgctgcagggc 3930
Db 4318 CATCTCGAAGCGGCTTGGCCGAGGGCGTCTGGCTGCTATTCGCGCGCCGATGTA 4377
QY 3931 tgtcggaacctgcaaggctcgctgctgcagcgagaccgatcccgcgatgcgaacc 3990
Db 4378 CCTGCGCAACCTGCAAGGTGCAAGTGTGAAGCGAGGCGCCACACGCGCGCTCAC 4437
QY 3991 cgttcgctgtagtatttcgagcgaggaaggaagcagcgctcgtgctgctgcgacgc 4050
Db 4438 CCTTTGCCCTGATGGACATGGAGCGTGAAGAGGCAAGGCTCTGGCTGCTGCGCCATTC 4497
QY 4051 tgcagccgaacacctgatcagggccgagctgcagagagagcgagatcggaatcatcc 4110
Db 4498 CCATGAGCGATATGTTAGAGGCGGATATCGAGTGTGATCGGATTCGCGCGCCATC 4557
QY 4111 cggtcagggaacttcggccgagctcaagcgatcgacagctcagcgacacatcaagt 4170
Db 4558 AGCTCGAGGACTACCGCGGGTGGTTCAGCGCCCTGCTGACCTGTGCGCGGACCATCAAGG 4617
QY 4171 cgatccgctgaagtgtcgcagcgatcccgcttcaggcgggccagtcagtgagctcg 4230
Db 4618 GTGTGCATCAAGCTCATCGGCGATGACCTTCAGGCGCGGCAATATCATCAACCTGA 4677
QY 4231 agatcccgccctcgggcagagcgcggttctcgatcggaacgcgcccgcgaagctcg 4290
Db 4678 CCTCGCGGGCGTTGAAGGATCCCGCGCTTCTCCCTGGCCAAACCGCGGAGCCAGG--- 4734
QY 4291 cggccaccgcgagatcgaaactgacgctgcggcaggtcgggcggtcgcgacggct 4350
Db 4735 -----CCGACGAGTGGAGTTGCATATCCGCTGCTGAGGGCGCGCGCCACCACT 4788
QY 4351 acctgcagagcaactggcgagcgggcgagcgtgctgcgctgcggcgccgttacggcgct 4410

Db 4789 TCATCCACCGCAGCTCAAGGTGCGGATGCGGTTCGAGCTGTCGGGCGCTCAGGSCACT 4848
QY 4411 tcttctgctgctgcctcgccgagcgccgagctatcttcaatgaggggttcggggtct 4470
Db 4849 TCTTCTGCTCGGTGATTCCTCCAGCGCGGACCTGATTTTCATCCGCGCGCTCCGCTCTGT 4908
QY 4471 cgaagcccgctcgcgtatgcgagccctgctgcgaagcgcgctcaccgcgcgatacagc 4530
Db 4909 CCAGCCCGAGTTCGATGATCTTCGACCTGTTTCCAGCGCGGCGACAGCGGAGATCACCC 4968
QY 4531 tgggttaoagtcgagcgagcgcgagggagctctatacaacagacgaattcccgcgctgg 4590
Db 4969 TGTTCAGGGCGCGGCAACCGCGGAGCTTTTACAAACCGGAGCTGTTTGAAGAACTGG 5028
QY 4591 csgaaacccatccgaacttcaagtaagtcgagcgcgctgctcgaagcgcaacgcgcgcg 4650
Db 5029 CCGCGCGCCACTCCAACTTCAGTACGTGCGCGGCACTCAACAGGCCCCAGCAGACCCGG 5088
QY 4651 gggcgagcgtcgcaaggggttcgtgcagcagctgcgaagcacatttccggcgcgact 4710
Db 5089 AGTGGCAGGGCTTCAAGGCTTTTCCAGGACGCGCGCAAGGCCCACTTCGAGGGTCTGT 5148
QY 4711 tctcgggcccagcgctacgtgctgagggcgcccgcgatgacgagcggtgcatacaga 4770
Db 5149 TCAGTGGCGCACAAAGCTATCTCTCGGTTCGCGCGCGATGATCGACGCGGCAACCA 5208
QY 4771 cgtcatcagggcgccgctgttcgagcgcgacatctacgagaagttcatctcgcgcg 4830
Db 5209 CCTGATGACGAGGTTCGGCTGTTTCGAGCGGGACATCTTTATGGAGCGCTTTTTCACCGCG 5268
QY 4831 ccgacgcgcacagacgcgcgagcc-----cgctgttcccgcggtgtgacatg-gagcg 4883
Db 5269 CGGATGGGGCGGACGACGACCCGTTTCGGCTTGTTCAGGCGATCTCAGTGGCTCAT 5328
QY 4884 gggcgcgctatgctgggagcgctacgagtcgagcagaccgagcgctatgctgctgctc 4943
Db 5329 GAGCAGCCCCCGTTTCAGGTGACGAAACCAACAGCGCGGAGTCTTACCTTACCTCGCGCC 5388
QY 4944 cggcgagctgcgtgcgcgagcgcgaactcgcaactcgccgcgcgcatccgctcgctg 5003
Db 5389 GGATCAGTCCGGTCTGCGGCGATGGAGGAGCAGGGCAAGCGTTCGCTGCGCGCTGGT 5448
QY 5004 cctgaacgcgggtgcgctgctgcaagtgctgcgctgctgcgctgctgcgcaagt 5063
Db 5449 CCGTGGTGGTGGCTGTGGTGTGCAAGTGGCGTGTCTCAGCGCGACTAC---AGTG 5505
QY 5064 cggcgcatcagcgctgcccagtgagcgcggaagaagaagacgagcgctacgctgctg 5123
Db 5506 CGCGAGGATGAGTGCAGTTCAGGTTCCACCGGAGGCGCGGAGCAGGCGCTGGCTTGGC 5565
QY 5124 gtccgcgctgcgagcgagcgagcgtcgaaactcga 5160
Db 5566 CTGTCAACTGATCCACGCGCTGATCTGTATCATCGAA 5602

RESULT 11
AF012632/c

LOCUS AF012632 2934 bp DNA BCT 21-AUG-1997
DEFINITION *Ralstonia pickettii* crpD, crpC, crpB, and crpA genes, complete cds.
ACCESSION AF012632
VERSION AF012632.1 GI:2338423
KEYWORDS
SOURCE *Ralstonia pickettii*.
ORGANISM *Ralstonia pickettii*
Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
Ralstonia.
REFERENCE 1 (bases 1 to 2934)
AUTHORS Olsen R.H., Kukor J.J., Byrne, A.M. and Johnson, G.R.
TITLE Evidence for the evolution of a single component phenol/cresol
hydroxylase from a multicomponent toluene monooxygenase
JOURNAL J. Ind. Microbiol. (1997) in press
REFERENCE 2 (bases 1 to 2934)
AUTHORS Olsen R.H., Kukor, J.J., Byrne, A.M. and Johnson, G.R.

TITLE Direct Submission
JOURNAL Submitted (06-JUL-1997) Microbiology, University of Michigan, 5608
Med. Sci. II, Ann Arbor, MI 48109-0620, USA
FEATURES
source
/organism="Ralstonia pickettii"
/strain="PK01"
/db_xref="taxon:329"
complement(374..883)
/gene="crpD"
complement(374..883)
/gene="crpD"
/codon_start=1
/transl_table=11
/protein_id="AAB67108.1"
/db_xref="GI:2338427"
translation="MTRVAKKLGLKERYAAMRGLGWETTYQPMKRVFPYDTYEGI
KTHMDKWDPPRLTMDAYWKYQGEKEKLYAVIEAFQNNNGOLGSDARYNNALKLF
IQVTRWTRAPWVRPRAQFTGEGPWRMQSIDELRHYQTETHAISHYKNKYFNGMHS
PNHWDRVW"
complement(929..1198)
/gene="crpC"
complement(929..1198)
/gene="crpC"
/codon_start=1
/transl_table=11
/protein_id="AAB67107.1"
/db_xref="GI:2338426"
translation="MSQVFIQAFQANESRPVIEAIYTDNPEAVVYPTGLVKIDAPGR
LITRRTIEQTRPDLQQLVNLVTLSGHIDEDDDQLTLSNQH"
complement(1234..2241)
/gene="crpB"
complement(1234..2241)
/gene="crpB"
/codon_start=1
/transl_table=11
/protein_id="AAB67106.1"
/db_xref="GI:2338425"
translation="MNIDQAREITPLROTFAHVARLIGEGKAATRYQATYGAQMSA
NFHYRTWPAPHELYDAGRSRIKLADQKLPQROFYATWTRARQEAIVANFQF
VESRGLASGADLRALQVLLPLRHWAGNMNNASICAYGYTAFTAPAMFHAMD
DIGIAQLIRLATLDEPAVLEAGKLAWLEDPRWGLRRYVEDSFYVQDPVFLFAQN
LALDGLLYPLTYGHFVDDHLAQGGTAVAILTSPMPWHDETARWIDGVIKTMANGGE
PRAAGNRALLSQWFSQWADRAQAALAPVAEMALGEHGAALGEVRAALDAATKLGLS
L"
complement(2297..2506)
/gene="crpA"
complement(2297..2506)
/gene="crpA"
/codon_start=1
/transl_table=11
/protein_id="AAB67105.1"
/db_xref="GI:2338424"
translation="MTIQLSQSNEMPVVDLTRKYVSRIERRTDGLVSPFAIGWPEL
SYDVLLEPFAFACRHRHOVTRLGS"
complement(2577..2591)
complement(2720..2768)
/bound_moiety="fbut"
BASE COUNT 519 a 890 c 953 g 572 t
ORIGIN
Query Match 20.88; Score 1213.4; DB 2; Length 2934;
Best Local Similarity 72.0%; Pred. No. 1.2e-115;
Matches 1801; Conservative 0; Mismatches 636; Indels 66; Gaps 14;
QY 240 cagcaccacccagctcttcccggttcgacccgagcgaagtcggtcccggtgaccgcg 299
Db 2488 CAGCCCTCCAACGAGATGCCCGTGTACCTCAGCGCGAAGTACGTGAGTCGATCG 2429
QY 300 acgaacgcgcgcgcgttcctcgaattcgaattcgaattcgaattcgaattcgaattcga 359
Db 2428 CGGCGCACCGATGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2372

QY 360 gtccgagtcgagctgtgtctctctgcccgcgttcgagtcggttctgcccgcgaacagcaggtcacg 419
Db 2371 GTCCAGCTGTGTGTCGCCAGAGCCGCGCATTCACCGCGTTCTGTGAGCGCCACCGAGTGACG 2312
QY 420 cggctc-----gagctcgaagcgaacccatgacctgagagagcaagaag----- 463
Db 2311 AGGCTCGGTTCCYTGAGACGGCGAGACCGTCCATCGCTTCATCCATAAACCACCCAGGA 2252
QY 464 -----tgaccatcgagctgaagacagtcgacatcaagccgctccggtccacacctt 512
Db 2251 GACAACCCGCATGAATATCGACCTGCGAGGCCGCGAGATCAGCGCGCTGCGGAGACCTT 2192
QY 513 tgcgcatgtcgcgcagacaacatcggcgcgacaagaacgagcgcgtaccaggaagcat 572
Db 2191 CGGCCAGTTGCCCGCCTCATCGGTGAGGCAAGCGCGCCACCGTTTACCAGGAGCGCAC 2132
QY 573 gatggcgcgacgccccagagaaactccattaccgcccacacctgggaccgcgactacga 632
Db 2131 GTACGGCGCGCATGTCATGGCGAACTTTCATATCGCCCGACGTGGAGACCCCGCGCAGCA 2072
QY 633 gatcttcgataccgtcgcgctcggcgatcccgatggcgaaactgggtacgctgttgaggatcc 692
Db 2071 GTTGTACGACGCGCGCGCAGCGCATCCCGCTCGCGCACTGGCGCAAGCCCTGAAGGACCC 2012
QY 693 gcgccagttctactacgctcgtcgtggcgacacacgcgcgcgccagcagagatgcgacgga 752
Db 2011 CGGTCACTTATTACGCCCACTTGGACCATGACCGCGCGCCGCGCAGGAGCGGTGGA 1952
QY 753 gtcaacttcgagttcgtcaaatcgcgcgagatgatcgctgatcgccgacgacgtggc 812
Db 1951 AGCGAACTTCCAGTTTGTGAGTCCGCGGCGTGGGATCTGCGATGGCGGATGGCGTGG 1892
QY 813 cgcgcggcgctcgacgtgctggtgcgctgcgcacgcgcgctggtggcgcgacaacatgaa 872
Db 1891 CGACCGCGCTGTCAGGTGTGCTGCGCACATGCGCGCACGTGCGCTGGCGCGCAATATGAA 1832
QY 873 caacgcgacatctgcgcgtcggcgatcgacgacgagtggttaccgcgcgcgcgagtggtccca 932
Db 1831 CAACGCGTCACTTGCCTGATGCGTACGGCAGCGCTTACCGCACCCCGCATGTTCCA 1772
QY 933 tgcgtaggacaacctcgcgcgtcgcgcaataacctcacgctcgtcgcgtcgcgacgtggcgcga 992
Db 1771 TGCCATGGAGCATCTTGGGATCGCGCAGCTGATCACCGCGCTGCGCTGAGCTGGATGA 1712
QY 993 gcccgacgtgctggagcgcgccaaaggcgacctggaacccgcgcgcgcgcctggcgcgcgt 1052
Db 1711 GCGCGCGTGTGGAGCGCGCAAGCTCGCTGGCTGGAGATCCGCGCTGGCAAGGCTT 1652
QY 1053 gcgcgctacgtcgaggaacacgctggtcgtcgcgcgacccggtcgagctgttcacgcgca 1112
Db 1651 CGGTGCTAGCTTGAAGACAGCTTCGTGGTTCAGGACCCGCTCGAGCTGTCTGCGGCGCA 1592
QY 1113 gaacctcgcgtcgacgcctcgtctatccgctcgtctacgacccgctcgtcgacgacg 1172
Db 1591 GAACCTTGCATCGACGCGCTGTGTATCCGCTCTACGCTCAGCTTCGCTCGATGACCA 1532
QY 1173 gatcgctcgaaggcggtcgcgagtcgcgagtcgacgcgcttcacgcccgaattgca 1232
Db 1531 CCTCGCGCTGCAAGCGGGACGCGCGTGGCCATCTCACCAGCTTTCATGCCGAGTGCA 1472
QY 1233 caccgagtcgaacccgctgagtcgacgagcggtcgtggaagacgagatggc-----cgc 1280
Db 1471 CGAGCAAAACCGCGCTGGATCGACGCGCTCATCAAGACCATGGCGCAACGCGCGGAGCC 1412
QY 1281 cgaatccgacgaacacccgcgctcgtccgcgcgtcgcacacgcgacgtgtccgcgcgcgc 1340
Db 1411 CGAAGCTGCGCGTAAACCGCGCGTGTCTTCGAGTGGTTCGCGAGTGGCGGACCGCGC 1352
QY 1341 cgaggcgacactggcacccggtgcgcgacgcgcgctcagagatgccggggcgcgcgcgct 1400
Db 1351 CCAGGCTGCTGGCGCGGCTTCCGAGATGGCGTGGCGGAGACGACGCGCGCGCGGCTT 1292

D85083	5552 bp	DNA	BCT	11-NOV-1997
Acinetobacter sp.	gene for DMS oxygenase components, complete cds.			
D85083	D85083.1	GI:2605607		
DMS oxygenase component.				
Acinetobacter sp. (strain:20B) DNA.				
Acinetobacter sp.				
Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;				
Moraxellaceae; Acinetobacter.				
1 (bases 1 to 5552)				
Omori,T.				
Direct Submission				
Submitted (07-MAY-1996) to the DBJ/EMBL/GenBank databases. Toshio				
Omori, The University of Tokyo, Biotechnology Research Center;				
Yayoi 1-1-1, Bunkyo-ku, Tokyo 113, Japan				
(E-mail:asegyoehongo.ecc.u-tokyo.ac.jp, Tel:03-3812-2111(ex.3067),				
Fax:03-5802-3326)				
2 (sites)				
Horinouchi,M., Kasuga,K., Nojiri,H., Yamane,H. and Omori,T.				
Cloning and characterization of genes encoding an enzyme which				
oxidizes dimethyl sulfide in Acinetobacter sp. strain 20B				
FEMS Microbiol. Lett. 155 (1), 99-105 (1997)				
98005684				
Location/Qualifiers				
1..5552				
/organism="Acinetobacter sp."				
/strain="20B"				
/db_xref="taxon:472"				
75..365				
/gene="dsoA"				
75..365				
/gene="dsoA"				
/codon_start=1				
/transl_table=11				
/product="DMS oxygenase component"				
/protein_id="BAA23330.1"				
/db_xref="GI:2605608"				
/translation="MIDAKOPTALVKYIRITGERNAKFVEFDFAIQDPTLFLVELLPQ				
QAFOHCEINHVIEMTAEQQAQWDAQEDKRWYIEPTVLNHRQHSDDQQA"				
384..1385				
/gene="dsob"				
384..1385				
/gene="dsob"				
/codon_start=1				
/transl_table=11				
/product="DMS oxygenase component"				
/protein_id="BAA23331.1"				
/db_xref="GI:2605609"				
/translation="MTLEIKTSNLEPIRQTYVYIERFRGAKPATRYOEVSFDIOASTN				
FHYRLWPKDKTLNDKTHALQMDWYAFKDPQFYGYAVYVHRARLQDTAESHYAF				
EKRLVNNLSDEVKQKIIQRLPLPRYVEQANLHMSSGSAYGTVITQACIFAMDR				
LGMAYISRIGLLDNGTESLQQAQHWLNDETWPRLKLCESLSETEQDWFKLYILQ				
NLIIDSMQLBELVFGQDEWLVSAGGRDIAILTBFMKDCLTDLAKWSVSLKTAISESE				
DNKTLIQSWITELLPVQKQAFSAWAQTALTDSGIDSLINKISERSKAGTILLDLAA"				
1398..1667				
/gene="dsoc"				
1398..1667				
/gene="dsoc"				
/codon_start=1				
/transl_table=11				
/product="DMS oxygenase component"				
/protein_id="BAA23332.1"				
/db_xref="GI:2605610"				
/translation="MTSKVYLAIQNDNTSRYYIEAIEDQNPQATIQKLPAMIRVESTG				
ELVVRAETVSEKLGQNWQDIQELQNLMTILGQNVDEDDSDFTLKN"				
1726..3261				
/gene="dsod"				
1726..3261				
/gene="dsod"				
/codon_start=1				
/transl_table=11				
/product="DMS oxygenase component"				
/protein_id="BAA23333.1"				

RESULT 13
D85083

[illegible]

[illegible]


```
Db 3993 GAACGTGAAGAAATAAGTGTGGCTGTGTTGTCACCAAGCAATCTGACATGGTCTATC 4052
QY 4071 gaggcgacgtcgacgaggagccggatgctgaggaataatcccggtcagggacttcgggcc 4130
Db 4053 GAAGCTGATGTGATGAAGACGAGGATTTCTAGGCTATTAGTCCAAAGACTACCAAGCT 4112
QY 4131 gacgtcacgcgatgaacagctcacgcgaccataaagtcgactccctcgaagctgctg 4190
Db 4113 AAGTTATTGAATTAACGGATTTATCCCAACAGTTAAAGGTGTCCGACTTCAACTTGAT 4172
QY 4191 cagccgatccgtccaggcgccagtcagtcgacgtcgagatcccgccctcgggcag 4250
Db 4173 CGACCAATGCAGTTTCAAGCAGGCCAATACATCAATATCAACTCCCAAAATATTGAAGGA 4232
QY 4251 agccgcgcttcgatcgcaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4310
Db 4233 ACCGGCGCATTTTCAATCGCAATATACACCGAGCGA-----CAAGAATCTGATCGAA 4283
QY 4311 ctgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4370
Db 4284 TTACATATTCGTAAGTGCAAGGTGTGTGACAGCAACACGCTATGTGCAATGATGAACCTGAT 4343
QY 4371 acggcgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4430
Db 4344 GTTGGTGAAGAAATGGCGCTTTCAGGACCTTATGGACAATTTTGTGCGCTAAGTCTGAT 4403
QY 4431 gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4490
Db 4404 CAGCAAAATGTCATTTTCAATGCGAGGTGTGTGACAGCAACACGCTTTCAGGCGCTTCAATCGATGAT 4463
QY 4491 gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4550
Db 4464 CTGATCTTTTGAACACGCTGACCCCGTATCATTTATTTGTTTCCAAAGGTGCGCGGAT 4523
QY 4551 gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4610
Db 4524 GTTGTGTAATATATAACCGTGAATAATTCGAACAGCTTGTCAAGAAATATCCAAATTTTC 4583
QY 4611 acgtacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4670
Db 4584 AGATACATCCCTGCCTTAAATGCGCCTTAAGCCAGAACCAATGGACAGGTTTACAGGT 4643
QY 4671 ttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4730
Db 4644 TATGTACATCAACGGGTGTCATTTATTTGAAAATAAATGCAAGTGGGCAACAAAGCATAT 4703
QY 4731 ctgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4790
Db 4704 TTATGTGGCCCGCCACCAATGATGTATGTCGCCCATCTCCACCTTGATGCAAAAGCGTTTA 4763
QY 4791 ttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4836
Db 4764 TTTGAAAAGACATCCATACCGAAGCCTTTTAAAGTGTCTGCAGATG 4809
```

RESULT 15

```
AB017631
LOCUS      1154 bp      DNA      BCT      20-FEB-1999
DEFINITION Comamonas testosteroni gene for phenol hydroxylase alpha subunit,
            partial cds.
ACCESSION AB017631
VERSION   AB017631.1 GI:4240017
KEYWORDS phenol hydroxylase alpha subunit.
SOURCE    Comamonas testosteroni (strain:R5) DNA.
ORGANISM  Comamonas testosteroni
            Bacteria; Proteobacteria; beta subdivision; Comamonadaceae;
            Comamonas.
            1 (bases 1 to 1154)
REFERENCE  Futamata,H.
            Direct Submission
            Submitted (11-SEP-1998) to the DDBJ/EMBL/GenBank databases.
            Hiroyuki Futamata, Marine Biotechnology Institute, Microbiol
```

```
Consortia Research Group: 3-75-1 Heita, Kamaishi, Iwate 026-0001,
Japan (E-mail:hfutamata@kamaishi.mbio.co.jp, Tel:81-193-26-6544,
Fax:81-193-26-6592)
2 (sites)
Futamata,H., Watanabe,K. and Harayama,S.
Unique primary structure found in phenol hydroxylases exhibiting
high affinity towards trichloroethylene
Unpublished (1998)
JOURNAL
FEATURES
    source
        location/Qualifiers
            1..1154
                /organism="Comamonas testosteroni"
                /strain="R5"
                /db_xref="taxon:285"
                <1..>1154
                /codon_start=2
                /transl_table=11
                /product="phenol hydroxylase alpha subunit"
                /protein_id="BAA74796.1"
                /db_xref="GI:4240018"
                /translation="RLTMDAYWYKYGEGEKKLYAVIEAFQNNQGLGVTDARYLNALK
                LFTQGVVPLEYAHRGFAHAGRHFTGAGARVAAMQMSIDELRHFOETHALSNKYF
                NGHSSSHMFDVRWFUSVPKSFEDLISAGPFLEFVAVSFSEIVLTLNLFVPMFMSGA
                AHNGDLSTVTFGSAQSDSRHMTLGIETIKFMLEQDPGNVPIQVRWIDKFWRGYRV
                LTLVAMQDYMFLPKRVMSKWEAMWEAEENGALFRDLARYGIREPAPWKLACEGDKH
                ISHOANNIFYNTAAAPFHTWVPGEOEMWLSKYPDSDPKHYRPLEHYLAEQQAGK
                RFYSKTLPLMLCTTCOIPMGTEPGDATKICYRESDEGSKYHFCSDGCKH"
BASE COUNT      236 a      360 c      339 g      219 t
ORIGIN
```

```
Query Match      12.7%; Score 740.8; DB 1; Length 1154;
Best Local Similarity 77.7%; Pred. No. 2.3e-67;
Matches 895; Conservative 0; Mismatches 257; Indels 0; Gaps 0;

QY 1970 ccgcctgacgatgatgcgtactggaataccaggcgagaaagaaagctgtacgc 2029
Db 1 CCGCCTGACCATGACGCCCTACTGGAAGTACCAGGCGCAAAAGAGAAAAGCTCTACGC 60
QY 2030 ggtgatcgacgcgttcacgagaacacgcgttccctcgcgtagcagaccccgctacat 2089
Db 61 CGTGATCGAGGCGCTTTGCCAGAGAACACGCCCTAGTTGGCGGTGACCATCGCGCTATCT 120
QY 2090 caacgcgctgaagctgttccctccagcgctgacccgctcgaatcacctcgcaaccgcg 2149
Db 121 CAACGCGCTCAAGCTGTTCATTACAGGCGGTGTCGCCCTGGAGTATTACGCCCAACCGCG 180
QY 2150 ctctgcgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2209
Db 181 TTTTGGCCATCGCGGCGCGCATTTTCACCGGTGCGGGGCGCACGCGTGGCGGCGAGATGCA 240
QY 2210 gtcgatcgacgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2269
Db 241 GTCCATCGACGAGTGCGCCACTTCCAGACCGAGACCCATCGCGCTGTCCAACATACACAA 300
QY 2270 gtcttcaacgcggttccatcactcgaaccagtggttcgcgcgcgcgcgcgcgcgcgcgcgcgc 2329
Db 301 GTACTTCAACGCATGACACGCTCCAGCCACTGGTTCCACCGGTCTGTTTCTGTGCGGT 360
QY 2330 gcggaagctcttcttcgagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2389
Db 361 GCCCAAGTCTCTTTTCAAGATGCGCTGAGCGAGCGGCCGCTTCGAGTTCTCTACGGCAGT 420
QY 2390 cagctctcttcgaatacactgtcgaacacctgtcttgcgtgcgcgcgcgcgcgcgcgcgcgcgc 2449
Db 421 GAGCTTCTCTCGAATATGTGTGACCAATCTGCTGTCTGTCGCCCTTCATGTGCGGCGC 480
QY 2450 cgctcaacgcggtgacatgctcgcaccgtcagcttcccttcgcgcgcgcgcgcgcgcgcgcgcgcgc 2509
Db 481 GGCGCAACACGGCGATCTGTCCACCGTGACCTTTGGCTTCTCGGCGCAAGGATGAGTC 540
QY 2510 gcgtcacatgacgcctcgccatcgatgcataagttcgtcgaacagaccgcgcgcgcgcgcgcgcgcgc 2569
Db 541 TCGCCATATGACGCTGGGCATCGAATGCATCAAGTTTCATGCTGAGCAGGACCGCGGCA 600
```

```
Qy 2570 cgtgccgacgtgcagcgctggtacacaaagtggtttctggcgcgtaccgcgctgctgac 2629
Db 601 TGTGCCCATCTGCAGCGCTGGATCGCAAAATGGTTCTGGCGCGGCTACCGCGTCTGTGAC 660
Qy 2630 gctggctgcgatgatgactacatgcagcccaagcgcgtgatgagctggcgcgagtc 2689
Db 661 CCTGTGGCCATGATGCAAGGATTACATGCTGCCAAAGCGCTGATGAGCTGGAAGGAAGC 720
Qy 2690 gtggaagatgtacccagcagaagcgcgcgcgcgtgttcaagtatctcgcgcgctacgg 2749
Db 721 CTGGAGATGATGCCGAGGAAACGGCGCGCTCTGTTCGCGACCTGGCGCGCTACGG 780
Qy 2750 catctgcgagccgaagggtggtgcagcgacctgcgaagcgaaggtcacatcagccacca 2809
Db 781 CATTCGCGAGCCCGCAGCGCTGGAAGCTGGCCCTGCGAAGGTAAAGCACACATCAGCCACCA 840
Qy 2810 ggcgtggtcgcagcttctacgcttcaacgcgcctcgcgcgttcacacacctgggtgcgcgac 2869
Db 841 GGCCTGGAACATCTTCTACAACTATACGGCGCGCGCACCTTTTCACACCTGGGTGCCCGG 900
Qy 2870 cgaagacgaatggcgtgctgcggcggaagtatcccgactcgttcgacgcgtactaccg 2929
Db 901 CGAGCAGGAGATGCAGTGGCTGTCCGAAATAACCCGACACAGCTTTGACAAAGCACTACCG 960
Qy 2930 cccgcgcttcgactcactcgtgggcgcagcgccagggcgcggcgaacgcgttctacatgaagac 2989
Db 961 TCCCGCGCTGGAGCACTACCTTTGCCGAGCAGCAGCGCGGCAAGCGCTTTTACAGCAAGAC 1020
Qy 2990 gctgccgatcgtgtgccagacgtgccagatcccgcgatcgtttcacccgagccgcggcaaccc 3049
Db 1021 CCTGCCCATGCTGTGCACCACTGTCAGATCCCTATGGGCTTACCGGAGCCCGGTGACGC 1080
Qy 3050 gacgaagatcgcgcgcgcgaatcgaactactcgcgcgaacaaagtccacttctgcagcga 3109
Db 1081 CACCAAGATCTGTATCGCGAGTCCGACTAGAGGGCAGCAAAATACCACTTCTGCAAGCA 1140
Qy 3110 ccaactgcaagga 3121
Db 1141 CGGCTGCANAGCA 1152
```

Search completed: September 26, 2000, 19:18:59
Job time: 12687 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2000, 15:50:55 ; Search time 153.15 Seconds
(without alignments)
9520.858 Million cell updates/sec

Title: US-09-430-029-1
Perfect score: 5828
Sequence: 1 gatcatttcacaaatgcgc.....tcggcccggaacacgacgc 5828

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues
Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2490.6	42.7	9785	1 T44457	DNA encoding tolue
2	2072.8	35.6	12019	1 X18867	Alcaligenes sp. Po
3	1854.8	31.8	4800	1 X05350	Phenol hydroxylase
4	1854.8	31.8	4800	1 X01590	Pseudomonas putida
5	327.2	5.6	652	1 X04453	Strain M07 oxygena
6	201	3.4	6379	1 Q79569	Nocardia coralina
7	201	3.4	6379	1 T17418	Nocardia coralina
8	137.2	2.4	9785	1 T44457	DNA encoding tolue
9	109.8	1.9	1240	1 N60800	Portion of plasmid
10	109.8	1.9	1992	1 N60855	Sequence encoding
11	109.8	1.9	2004	1 N60911	Sequence encoding
12	109.6	1.9	924	1 Q04017	Metapyrocatechase
13	100.2	1.7	958	1 N30167	Sequence of gene x
14	94.4	1.6	114955	1 X53491	Human adenosine A1
15	88.6	1.5	114955	1 X53491	Human adenosine A1
16	73.8	1.3	1266	1 X25773	S. erythraea erythr
17	73.8	1.3	3412	1 X25772	S. erythraea erythr
18	72.4	1.2	44377	1 T78508	Platenolide synth
19	72.4	1.2	44377	1 T80414	Sugar biosynthesis
20	72.2	1.2	3756	1 T72684	Sugar biosynthesis
21	71	1.2	53789	1 V21187	Amicolatopsis medi
22	69.4	1.2	29879	1 Q46806	eryA region of S.
23	68.4	1.2	5676	1 V21186	Amicolatopsis medi
24	61	1.0	23666	1 Q10190	Cephalosporin anti
25	60.8	1.0	8169	1 V26609	Actinomadura hibis
26	60.6	1.0	1224	1 V60558	Gtfid gene from Amy
27	59.8	1.0	15872	1 T68715	Streptomyces venez
28	59.4	1.0	2634	1 V22334	Microbispora therm
29	59.2	1.0	30001	1 T61016	Total DNA sequence
30	59.2	1.0	30001	1 X05110	S. aureofaciens DN
31	59	1.0	4020	1 T91361	Orf virus genomic
32	58.8	1.0	1204	1 V69560	Soil derived pepti
33	58.2	1.0	1227	1 V84121	Pentaclethra macro

ALIGNMENTS

RESULT 1

T44457	ID	T44457 standard; DNA; 9785 BP.			
AC	T44457:				
DT	29-JAN-1997	(first entry)			
DE	DNA encoding toluene ortho-monoxygenase.				
KW	tom; pTOM; self-transmissible; constitutive; bioindicator; pollutant;				
KW	breakdown; trichloroethylene; TCE; degradation; ss.				
OS	Pseudomonas cepacia strain PRL-23.				
FH	Key	Location/Qualifiers			
FT	cds	2647..2757			
FT		/*tag= a			
FT		/product= tomA0			
FT		/note= "no start or stop codons included; encodes			
FT		W06805"			
FT	cds	2872..3866			
FT		/*tag= b			
FT		/product= tomA1			
FT		/note= "encodes W06800"			
FT	misc_difference	3862..3866			
FT		/*tag= c			
FT		/note= "apparent deletion of nucleotide; encodes			
FT		cysteine and stop"			
FT	cds	3900..4175			
FT		/*tag= d			
FT		/product= tomA2			
FT		/transl_except= (4173..4175, aa:Glx)			
FT		/note= "encodes W06801"			
FT	cds	4198..5753			
FT		/*tag= e			
FT		/product= tomA3			
FT		/note= "encodes W06802"			
FT	misc_difference	5116..5117			
FT		/*tag= f			
FT		/transl_except= (5116..5117, aa:Ala)			
FT		/note= "apparent deletion of nucleotide"			
FT	misc_difference	5133..5135			
FT		/*tag= g			
FT		/note= "encodes Ala-Arg; apparent deletion of 3			
FT		nucleotides"			
FT	cds	5750..6030			
FT		/*tag= h			
FT		/product= tomA4			
FT	misc_difference	6026..6027			
FT		/*tag= i			
FT		/transl_except= (6026..6027, aa:Ala)			
FT		/note= "apparent deletion of nucleotide; encodes			
FT		W06803"			
FT	cds	6185..7245			
FT		/*tag= j			
FT		/product= tomA5			
FT		/note= "encodes W06804"			
FT	misc_difference	6224..6225			
FT		/*tag= k			
FT		/note= "apparent insertion"			
FT	misc_difference	6397			

Infected cell prot
The nucleotide seq
Sorghum bicolor (L
Streptomyces prist
GtfC gene from Amy
snBR gene encoding
S. clavuligerus cl
Tylactone synthase
Sequence encoding
Neomycin phosphor
Maize nitrite redu
Hybrid srmg/tylg O

Qy	922	gcgatgttccatgcgattgacaacactcggcgtcgcgcgaataacctacacgcgtctcgcgctc	981
Db	890	TGCACCTACTTGC CGCATGGATCCCTTCGCATTCGCCAGTAGCTTCCGCATCGCGATCGGCATG	949
Qy	982	gcgatggcggagcccgac-----gtcctggagcgcgccaaggcgacctgaccccgac	1035
Db	950	CTGCTGGATGGCAACACCCGGTGTGGCGCTGGAGACCGCCAAAGGTGGCTTGGATGGAGAGC	1009
Qy	1036	gcgcctggcagcgcgtgcgcgcctacgtcagaggacaacgtggctgcgcgatacccgctc	1095
Db	1010	CGGCGTGGCAGCGATGCGCCGGTGTGCGAGCGCAGCTTTGTGATCGAGGACTGGTTC	1069
Qy	1096	gagctgttaatcgcgcgagaaactcgcgtcgcgcgctcgtctgctatccgcctcgtctacgac	1155
Db	1070	GAGACCTTTGTCAACCCAGAACCTGTGTCTGACGAGCCCTGCTTTACCCGCTGGTGTACAG	1129
Qy	1156	cgtctcgtcgaacaacgcgtcgtcgaaggcgactcgcgcagtcgagatcgaatcgaacgcg	1215
Db	1130	C---ACGCCAGCGGGCCATCTGTGCGCGCTCGGGCACCGGGCTGGCGGTGCTGACCGAG	1186
Qy	1216	ttcatgccggaattggcacaccgagtcgaaccgctggatcgcgcgcgctgtgaagaacgatg	1275
Db	1187	TTCATGAATGACTGGCGCGAGGACGACGACGCTGGGTGACGCGGTGATCCACACCGCC	1246
Qy	1276	gcgcgcgaatccgaagcaaacccgcgctgtctgcgcgcctggacacgcgactggtccgcg	1335
Db	1247	CGGCGAGAGTCGATGCCAATCGCATGTGTGTCTCGGCTGGGCGCGCGCGCGCGCG	1306
Qy	1336	cgcgcgaaggcgcactgcaccgctggcgcgcgcgcgcgtgcaggatgcgggcgcgcg	1395
Db	1307	CAGGTGGCTGAGCGCTGTGTGCGGTGGCGCCACACTTCTGGGCGAGGGCGGCGAGCAG	1366
Qy	1396	gcgctcgaagaaagtgcgcgagcagttccacgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1455
Db	1367	GTGGTGGCGCTGTGCTGGAGCAGTTCGACGTTGCGCTTTAAACAGCTTGGCCTTGGCGCC	1426
Qy	1456	tgcgcgcggaatcctcccttaaccacaaggaaatgccacatgctccaacgattatccatgcc	1515
Db	1427	TGAGGA-----AACCGGCATGACCGGCCAATGTGTATATCGCG	1463
Qy	1516	ttcaggccaatgaggactccagacgatcgtggatcgatcgtcgcgcgacaacccgcgc	1575
Db	1464	CTGCAGACCAACGACGACACCCGTCCTCATTCGAGGCCATCACCGAGGCCAACCCGCAC	1523
Qy	1576	gcggtggtggtcagtcgcgccggcatggtccaagatcgcgcgcgcgcgcgcgcgcgcgcgc	1635
Db	1524	CGGGTGGTGTGCGAGTTTCCCGCCATGTGTGAAGATCGATGCGCGGGACACCTGACCATC	1583
Qy	1636	cgccgcgaacgatcgaggaaactgaccgcgcgcgcgttcgcacctgcagcagctccaggtc	1695
Db	1584	CTCGCGAACTGGTCTCCGACAAGCTCGGGCGCGACTGGGACCTGCGAGGAGATTACCTTG	1643
Qy	1696	aacctgatcacgctgtcagggccacatcgacgagcagcagcaggttcaacgctgagctcg	1755
Db	1644	AACCTGATTTGCTGTCTGGGAAACAATCGACGGAAGACGACGAGCGCTTCAACCTGCACTGG	1703
Qy	1756	tgcactgaa-cgcgcgcgcacgcgcacgcgcacaacacgcggagacacgaattggacacgcc	1814
Db	1704	AACGCCATGAATCGGCGGCCACCGCAAAAGACAAACGAGAGACAAAACCATGGACGCAC	1763
Qy	1815	acgctcaaaaaaactcgtgaagcagcgtacgcgcgcaatgacgcgcgcgcgcgcgcgcgcgc	1874
Db	1764	GC-----AAGAAGCTGAACCTGCGGGAAAAAATACGGGACGATGACGCGCGACCTCGGC	1816
Qy	1875	tgggagcagcctaccgcgctggaacaagttctcccgctacgaccgctcagagggcac	1934
Db	1817	TGSGAAACCACTACGAGCCGATGACAAGGTCCTTCCTTACGACAAAGTACGAAGGCATC	1876
Qy	1935	aagatccacgactgggacaagtgggtcgacccgcttcgcgctcgcgtatgagatgggactcgg	1994
Db	1877	AAGATCCATGACTGGGACAAATTGGGAAGACGCGTTTTCGATGACCATGACGAGCTACTGG	1936

[illegible]


```

Db 3017 AGCTACACGCGATCGCGTTCCTCCACTTCTGCTCGAGCGCTGCAAGGACATCTTCGATGGC 3076
Qy 3135 gagccgcgaataacgtgcagcgctggtcgctgcgctgcacagatccatcagggcaactcgc 3194
Db 3077 GAGCGCGAAGAGTACGCGCAGCGCTGGTGTGCTGTGTCACACAGATCTACGAGGCAACTGC 3136
Qy 3195 ttccgcgcgatgaggacccggcgcgagggcttcgatccgcctgcgcgcgcgctgctgcac 3254
Db 3137 -----GGCGGTGATCGCTGAGCGAGCTGCTCAAG 3166
Qy 3255 tactacgcggtgacgatggcgcgacaaacctcgatttcgcgctgcgctggaagaccagaag 3314
Db 3167 TGGTACCGCATCAACCTTGGCGCGGACAACTGGATTGGAAGTTTCAGGTTCCAGGATCAGAA 3226
Qy 3315 aactcgcggtggtgcggtgcgagcccaacgcgcaactgaccgcgcaacagaacaatttt 3374
Db 3227 AACTGGAAAGCGCTGGAAGGCGTGCC----- 3252
Qy 3375 gacgagggccgcgaagcgcgatgcgcgaacgcggcgacagcgagagacaacatggcg 3434
Db 3253 -----GGGACAGCGTGCCTGAGCGCGCGCATACAAGAGATAAGGAGACCAACCATGTCCG 3307
Qy 3435 tcctcgcgtcaaaacctacgacttcccggtgaagatgcgctgcgagagtttccg---- 3491
Db 3308 TCGTATCCATTGGCCCTTACAGCTTGCAGCGCGCGATCGGGAGGCGCGTGTTCACAGGCA 3367
Qy 3492 cgcgcgtctctacgtgtgctgggaaacacatctgatttccccgcgcgcgtttctgcctgc 3551
Db 3368 ACCGCTGTCTATATCGCGCTGGACCGCCACCTGCTGCTTTGTGCGCGCATTTGCCCTTC 3427
Qy 3552 cgtgcgcgcgcgaatcgcgttgcgcgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3611
Db 3428 CTTTGCCTCCCTGATGCGCTGCGGATGCTGCTGGAGAACGTCGCTCGCGGCGTCTACG 3487
Qy 3612 gctatcaccccgacttcgcgaagatcgactgggacgcgctgcgagtggttccggtgcggcg 3671
Db 3488 GCTATCACCCCGACTTTGCCCGCATCGACTGGAGCGCGTGGAGTGGCTGCGCGCGCGCG 3547
Qy 3672 agcgcgtggcgccgagaccgcgaagagcctggccgcgcgaacgcgcctgcgggcacaaaggacc 3731
Db 3548 AGCCCTTGGCAGCCAGACCTTGACCGCACGCTGGAAGAGAACGCGCTGGCGCCACAAAGCGG 3607
Qy 3732 tgatcagcttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3791
Db 3608 TGATCCGCTTACGACACCGGGCTGGATGGCATTTGGCGGAGTGCGAGCTTAACAGTGGC 3667
Qy 3792 gcgacgagcgaaccatcatgagccaccaacttaacctgcgagccgctggcgctgcacgac 3851
Db 3668 AGCTAAGGTAGGAGACACAAGATGTATTCCTGACCATTTGAACCGATCGGGCAGACCATC 3727
Qy 3852 gaggtcaggaggaagacagatgctcgtatgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3911
Db 3728 CCCATCGCGCGCGGCGCAGACGCTGTGATGCTGCTGCCCTGCGCACGCGGCTGTGGCTGCCG 3787
Qy 3912 cagcgcgtctgcagcgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3971
Db 3788 CACGCTGCTGCCACGCGCTGTGCCACCTGCAAGGTGCGAGGTGGTGGAGGGCGAAATTC 3847
Qy 3972 gatcccgcgatgcggaaccccttcgcgtgatgatttcgcgagcgaggaagggcaagcg 4031
Db 3848 GAGCATGGAGAGGCTCCAGCTTCCGCTGTGATGGACTTCGAGCGCGACAGCGGGCAGTGC 3907
Qy 4032 ctgcgcgtctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4091
Db 3908 CTGGCTTGTCTGGCGCACCGCGCAGTCTCCGACATGTGTGATCGAGGCGGATATCGAGGAAC 3967
Qy 4092 ccgcatgcggaatacatctcccggtcaaggacttcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4151
Db 3968 GCCACTCGCTGGCGCTGCCCTGGCTGACTATCGTGGCGAGGTGGTGGAGCGCGCGCG 4027
Qy 4152 ctacgcgcgcacat-----caagtcgatccgcctgaagctgtgcgcgcgcgcgcgcgccttc 4205
```

```

Db 4028 CTGACCCCCACCATCCGCGGCACTCTGGCTGCGCGTGAAGGGGGGGCGCGCTCCCTTC 4087
Qy 4206 caggcggccagtagtgcagctcagatcccgagatcccgccctcgggcagagcgcgcgttctcg 4265
Db 4088 CAGCGCGCGCAGTACTCAACCTGCGCTGCGGGCTGCGACACCGCGTGCCTTCG 4147
Qy 4266 atcgcgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4325
Db 4148 CTGGCAACCGTCCCGCGGATGACCTGG-----TGGAGCTCATGTGCGGCGG 4195
Qy 4326 gtccggcgcggttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4385
Db 4196 GTGGAAGCGGCGGACGCTACCGGCTACCTGCACGATACGCTGCGGTGGTGACAACTC 4255
Qy 4386 cgcctcgtggcccgtagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4445
Db 4256 GGGTTTTCGCGGCTTACGCGCGCTTCTGCTGCGCAAGTCAGCGCAAAAGCCGATGTG 4315
Qy 4446 ttcatggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4505
Db 4316 TTCTTGGCGGCGGCTCGGGCTTGTCCAGCCCGCGGCAATGATCTTGACATGTGGCT 4375
Qy 4506 agcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4565
Db 4376 GCCGCGAGACCTTCCGATCAGCGTGTGCAAGCGCGCAACCGCAGGAGTGTAC 4435
Qy 4566 taccacgcgaattccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4625
Db 4436 TACGACGAGGCTTCCGTGCTGCTGCGCGCGCGCACCCCAACTTCCGCTATGTGCCCGCG 4495
Qy 4626 ctgtccgaagcgcaaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4685
Db 4496 CTCCTCGAGAACCGCGCGGACAGCGCTGGGACGCGCGCGCGGCTATGTGCATGACGTC 4555
Qy 4686 gcgaaggcacatttcgg-----cggcgacttccggcaccagcgctacctg 4733
Db 4556 CTGCACGCGCTTTACGCAATGGCGCGACCGCGCGACTTCCGTGGCCACAAAGSCCTATCTG 4615
Qy 4734 tgcggcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4793
Db 4616 TCGCGCGCGCGCGCGATGATCGAAGCTGATCCCGACGTTGATGACGGCGCGGTTC 4675
Qy 4794 gagcgcgcacatcatcacgaagaattcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4850
Db 4676 GAGGAGGACATCCACACGAGAAATTCATCTCGCGCGCGGACGACAGACAGCGCGCGC 4735
Qy 4851 agccgcgttc 4862
Db 4736 AGCCGCGTTC 4747

RESULT 4
X01590
ID X01590 standard; DNA; 4800 BP.
AC X01590;
DT 05-MAY-1999 (first entry)
DE Pseudomonas putida phe operon.
KW Chlorinated ethylene decomposition; phe2; pheA; pheB; pheC; pheD; pheE;
OS Pseudomonas putida.
FH Key Location/Qualifiers
FT CDS 127..345
FT /tag= a
FT /product= phe2
FT 434..1429
FT /tag= b
FT /product= pheA
FT 1440..1712
FT /tag= c
FT /product= pheB
FT 1754..3268
FT /tag= d
FT /product= pheC
```


CC metaprotease, bonded at its C-terminal to the N-terminal of
CC GM-CSF. Expression vectors pMG1 and pMG3 also contain the tac
CC promoter/operator and the SD sequence of metaprotease.
SQ Sequence 924 BP; 200 A; 266 C; 265 G; 193 T;

Query Match	1.98;	Score 109.6;	DB 1;	Length 924;
Best Local Similarity	51.7%;	Pred. No. 6.7e-11;		
Matches 275;	Conservative 0;	Mismatches 254;	Indels 3;	Gaps 1;
Qy 5245	tcacatggggtgatcgctattgctcatgtcagctcagctgaagtgatgacatgaagcgg	5304		
Db				
Db 2	TGACAAAGGTGTATGCGACGGCGCCATGTGCAGCTCGTGATCGACATGAGCAAGG	61		
Qy 5305	cgtcgctcattacgtaacgctgctcgcatcgaggaaacgatcgcgacgcgcgggca	5364		
Db				
Db 62	CCCTGGAACACTACGTCGAGTTGCTGGSCCTGTCGAGATGCAGCGTGACGACGAGGCC	121		
Qy 5365	acgtctactgaatgctggagcaaatgggacaagtattcgcctgatactgctgcgtcgg	5424		
Db				
Db 122	GTGCTATCTCAAGGCTTGGACCGAAGTGGATAAGTTTCCCTGGTCTACCGCAGGCTG	181		
Qy 5425	atcaggcggggctcaagcatcgccctacaaggtcgagacagcgcgatctggatgcgc	5484		
Db				
Db 182	ACGAGCCGGGATCGATTTATGCGTTTCAAGTTGTGGATGAGGATGCTCTCGGGCAAC	241		
Qy 5485	tgcagcagcgcatcgaagcgtacgggatacgcgaccgagatgctgcgcaaggcgcgctgc	5544		
Db				
Db 242	TGGAGCGGATCTGATGSCATATGCTGTGCGGTGACAGCTACCCGCGAGTGAACTGA	301		
Qy 5545	cggcggtcggcgccaaactcggttctctgctgcgagcggccataaactcggtcttgcg	5604		
Db				
Db 302	ACAGTTGTGGCGCGCGTCCAGGCGCTCCAGGCGCCCTCGGGCATCAC'TTCAGTTGTATG	361		
Qy 5605	cgaagaaggcgtggtgggcaaccgg---tcggctcgctgaaccccgatcgggtggccgg	5661		
Db				
Db 362	CAGACAAGGAATATACTTGGAAAGTGGGTTTGAATAGCGTCAATCCCGAGCATGGCGCG	421		
Qy 5662	acgacattccgggctcgggcgctgcactggtcgaccactgctcgtgatgtgcgaactga	5721		
Db				
Db 422	CGATCTCAAAGGTATGGCGCTGTGCGTTCGACCACGCCCTCATGTATGGCAGCAAT	481		
Qy 5722	accggaggccggcgctgaaccgcgtcaggagaaacacgcgttcattggccga	5773		
Db				
Db 482	TGCGGGGACCTATGACCTGTTTCAACAAAGTGCTCGGTTTCTATCTGCGCCA	533		

```

RESULT 13
N30167
ID N30167 standard; DNA; 958 BP.
AC N30167;
DT 31-MAY-1992 (first entry)
DE Sequence of gene xyle encoding catechol 2,3 oxygenase (C2,3-O)
DE and a ribosome binding site.
KW Catechol; food industry; pharmaceuticals; cosmetics; disinfectant;
KW ss.
FH Key
FT rbs
FT cds
FT Location/Qualifiers
FT 16..25
FT /*tag= a
FT 30..950
FT /*tag= b
PN EP-86139-A.
PD 17-AUG-1983.
PF 27-JAN-1983; 400187.
PR 01-FEB-1982; FR-001574.
PA (TRAN-) TRANSGENE SA.
PI Zakowski M, Garfney D, Speck D, Lecocq JP; .
PD WPI; 83-742602/34.
DR P-PSDB; p30266.
DR P-PSDB; p30266.
PT pLasmid vector for synthesis of catechol 2,3-oxygenase - carries
PT xyl E gene and promoter for expression in gram-positive bacteria,
PT esp. Bacillus subtilis
PT Example; Fig 3; 66pp; French.
PS

```

CC	C2,3-O has uses in the food industry, the pharmaceutical industry, the cosmetics industry and as a disinfectant. It converts catechols to aldehydes. The inventors claim a plasmid vector carrying at least the xyle gene and a promoter for the expression of the gene in a Gram positive bacteria, pref. Bacillus subtilis.	CC	Gram positive bacteria, pref. Bacillus subtilis.	CC	Sequence	588 BP;	211 A;	270 C;	276 G;	201 T;
Query Match	1.7%;	Score	100.2;	DB	1;	Length	958;			
Best Local Similarity	48.8%;	Pred. No.	2.7e-09;							
Matches	270;	Conservative	0;	Mismatches	283;	Indels	0;	Gaps	0;	
Qy	5221	atcaacaagtaacaccaggaggagactcaaccatgggtgtgatcgtattgtgtcatgtcagtc	5280							
Db	7	ATGAACATATGAAGAGGTGACGCTCATGAAACAAAGGTGTAAATGCGACCGGCCATGTGCAGC	66							
Qy	5281	tgaaggtgaatgagacatggaagcggcgctgcgtcattcaagtaacggtgcctggccatgcagg	5340							
Db	67	TGCGTGTACTGGCATGATGAGCAAGGCCCTGGAAACACTACGTCAGTGTGTGGCCCTGATCG	126							
Qy	5341	aaacgatgcgcgcgcggcgggcgaacgtcttaccctgaatatgctgggaacaaatgggacaagt	5400							
Db	127	AGATGGACCGTGCAGCACACGAGGCCGTGTCTATCTGAAGGCTTGGACCGAAGTGGATAGT	186							
Qy	5401	attcgtgtacctgtcgcgtccgatccagtcaggcggggctcaagcatgccctacaaggtcg	5460							
Db	187	TTTCCTCTGGTCTACGCGAGGCTACGAGCGGGCATGGATTTTATGGGTTTCAAGGTTG	246							
Qy	5461	agcagcagcgcgatctggatgcgtgcagcagcgcacatcgaaagctacgggatcgcgacgc	5520							
Db	247	TGGATGAGGATGCTCTCCGGCACTGGAGCGGGATCTGATGCATATGGCTGTGCCGTG	306							
Qy	5521	agatgctgcccaaggcgcgtgcgcggcgtcgccgcgcaactcgcggttccgtgctgcgcga	5580							
Db	307	AGCAGCTACCCGCGAGGTGAACGTGAACATTTGTGGCCGCGCGTGGTTCAGGCCCTCCG	366							
Qy	5581	gcggccatgaactgoggctgttcgcgaagaaggcgcgtggtggcaccgcggtcggtcgcgc	5640							
Db	367	GGCATCATCTTCGAGTGTGATGCAGACAAAGGAATATACTGGAAGTGGGGTTTGAATGACG	426							
Qy	5641	tgaaccgcgacctggccgcgaagacattccggctcgccgtgcgctgcactggctcgaccact	5700							
Db	427	TCAATCCCGAGGCATGGCCCGCGCATGTGAAGGTATGGCGGCTGTGGCTTTCGACCACG	486							
Qy	5701	gcctgctgatgtgcgaactgaaccgcgaggccgcgctgaacgcgcgtcgaggagaaacacgc	5760							
Db	487	CCCTCATGTATGGCGACGAATTGCCGGGACCTATGACCTGTTCCACCAAGGTGCTCGGTT	546							
Qy	5761	gcttcatggccga	5773							
Db	547	TCTATCTGGCGGA	559							

RESULT	14	
X53491/C		
ID	X53491 standard; DNA; 114955 BP.	
AC	X53491;	
DT	05-JUL-1999 (first entry)	
DE	Human adenosine A1 receptor antisense oligonucleotide fragment.	
KW	Antisense oligonucleotide; multiple target; antisense treatment;	
KW	impaired respiration; inflammation; lung disease;	
KW	pulmonary vasoconstriction; inflammation; allergic rhinitis;	
KW	acute asthma; allergy; asthma; impeded respiration;	
KW	respiratory distress syndrome; pain; cystic fibrosis;	
KW	pulmonary hypertension; pulmonary vasoconstriction; emphysema;	
KW	chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;	
KW	colon cancer; breast cancer; lung cancer; pancreatic cancer;	
KW	hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;	
KW	prostate cancer; ss.	
OS	Synthetic.	
PN	W09913886-Al.	
PD	25-MAR-1999.	
PF	17-SEP-1998; U19419.	

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2000, 15:21:14 : Search time 1948.25 Seconds
(without alignments)
13191.571 Million cell updates/sec

Title: US-09-430-029-1
Perfect score: 5828
Sequence: 1 gatcatttcataaatgcgc.....tcggcccgaggcaacacgac 5828

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues 10495684
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: em_est1.*
2: em_est2.*
3: em_est3.*
4: em_est4.*
5: em_est5.*
6: em_est6.*
7: em_est7.*
8: em_est8.*
9: em_est9.*
10: em_est10.*
11: em_est11.*
12: em_est12.*
13: em_est13.*
14: em_est14.*
15: em_est15.*
16: em_est16.*
17: em_est17.*
18: em_est18.*
19: em_est19.*
20: gb_est1.*
21: gb_est2.*
22: gb_est3.*
23: gb_est4.*
24: gb_est5.*
25: gb_est6.*
26: gb_est7.*
27: gb_est8.*
28: gb_est9.*
29: gb_est10.*
30: gb_est11.*
31: gb_est12.*
32: gb_est13.*
33: gb_est14.*
34: gb_est15.*
35: gb_est16.*
36: gb_est17.*
37: gb_est18.*
38: gb_est19.*
39: gb_est20.*
40: gb_est21.*
41: gb_est22.*
42: gb_est23.*
43: gb_est24.*

44: gb_est25.*
45: gb_est26.*
46: gb_est27.*
47: gb_est28.*
48: gb_est29.*
49: gb_est30.*
50: gb_est31.*
51: gb_est32.*
52: em_est20.*
53: em_est21.*
54: em_est22.*
55: em_est23.*
56: em_est24.*
57: em_est25.*
58: em_est26.*
59: gb_est33.*
60: gb_est34.*
61: gb_est35.*
62: gb_est36.*
63: gb_est37.*
64: gb_est38.*
65: em_est27.*
66: em_est28.*
67: em_est29.*
68: em_est30.*
69: gb_est39.*
70: gb_est40.*
71: gb_est41.*
72: gb_est42.*
73: gb_est43.*
74: gb_est44.*
75: em_est31.*
76: em_est32.*
77: em_est33.*
78: em_est34.*
79: gb_est45.*
80: gb_est46.*
81: gb_est47.*
82: em_est35.*
83: em_est36.*
84: em_est37.*
85: gb_est48.*
86: gb_est49.*
87: gb_est50.*
88: gb_est51.*
89: gb_est52.*
90: gb_est53.*
91: gb_est54.*
92: gb_est55.*
93: gb_gss1.*
94: gb_gss2.*
95: gb_gss3.*
96: gb_gss4.*
97: em_gss1.*
98: em_gss2.*
99: em_gss3.*
100: em_gss4.*
101: gb_gss5.*
102: gb_gss6.*
103: gb_gss7.*
104: gb_gss8.*
105: gb_gss9.*
106: em_gss5.*
107: em_gss6.*
108: em_gss7.*
109: em_gss8.*
110: em_gss9.*
111: em_gss10.*
112: em_gss11.*
113: gb_gss10.*
114: gb_gss11.*
115: em_gss12.*
116: gb_gss12.*

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:12:01 ; Search time 92.32 Seconds
(without alignments)
248.587 Million cell updates/sec

Title: US-09-430-029-3
 Perfect score: 1723
 Sequence: 1 MTIELKTVDIKPLRHTFAHV.....ALDEVREOFHARAALGIAL 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

```
Searched:      225878 seqs, 69334122 residues
Total number of hits satisfying chosen parameters:  225878
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database :
SPTRMBL12:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Query		Length	DB	ID	Description
			Match	%				
1	1139	66.1	335	2	Q30592	Q30592	burkholderi	
2	1101.5	63.9	330	2	Q92NP6	Q92NP6	comamonas t	
3	1021.5	59.3	336	2	Q52570	Q52570	pseudomonas	
4	832	48.3	331	2	Q84959	Q84959	raistonia s	
5	778.5	45.2	331	2	Q52171	Q52171	pseudomonas	
6	770.5	44.7	342	2	Q52162	Q52162	pseudomonas	
7	640.5	37.2	333	2	Q43979	Q43979	acinetobact	
8	640.5	37.2	333	2	Q32439	Q32439	acinetobact	
9	222.5	12.9	330	2	Q87802	Q87802	pseudomonas	
10	193	11.2	329	2	Q51943	Q51943	burkholderi	
11	190.5	11.1	332	2	Q69182	Q69182	alcaligenes	
12	179.5	10.4	343	2	Q53025	Q53025	nocardia co	
13	167.5	9.7	341	2	Q9ZET3	Q9ZET3	xanthobacte	
14	134	7.8	332	2	Q07072	Q07072	burkholderi	
15	121.5	7.1	395	2	Q06117	Q06117	methylolcyst	
16	105.5	6.1	1092	3	Q13462	Q13462	emericeila	
17	105.5	6.1	1092	3	P78622	P78622	emericeila	
18	100.5	5.8	567	11	Q9WU19	Q9WU19	mus musculus	
19	100.5	5.8	570	11	Q9WU01	Q9WU01	mus musculus	

ALIGNMENTS

RESULT	1
O30592	
ID	O30592
AC	O30592;
DT	01-JAN-1998
DT	01-JAN-1998
DT	01-JAN-1998
DT	01-JAN-1998
PRT:	PRELIMINARY; (TREMBlrel. 05, Created)
	(TREMBlrel. 05, Last sequence update)
	(TREMBlrel. 05, Last annotation update)

Query Match	66.1%	Score 1139;	DB 2;	Length 335;
Best Local Similarity	64.5%;	pred. No. 2.6e-86;		
Matches 216;	Conservative 41;	Mismatches 74;	Indels 4;	Gaps 1;
Qy	1	MTIELKTVDDIKPLRHTTAHVAONIGDGKTKATRYCEGMMGAQOQENHFYRPTWDPDYIEFD	60	
Db	1	MNIDLQAREITPLRQTAHVARLLGEGKATRYQEATYGAQSMANFHYRPTWDPAHLEYD	60	
Qy	61	PSRAIRMANWYALKDPRQFYASWATYRPAQQDAMESNFVEFSRRMIGLRRDDVAAARA	120	
Db	61	AGRSRIBLADQWALKDPRQFYATYTWTRARQQEAVANFQVFSRGSLASAMGDALRDRA	120	
Qy	121	LDVLVPLRHAAGMANMNAOICGALCYGVTFAPAMFHAMDNILGVAQYLTRLALAMAPDV	180	
Db	121	LQVLLPLRHYAWGANMNASICAYGYGTAFTAPAMFHAMDDLGIAQLITRALTLDEPAV	180	
Qy	181	LEAKATWTDRDAWQPLRRVVEDPLVVADPVELFTANLALDGLLYPLVYDRFVDRIAL	240	
Db	181	LEAGKLAWLEDPRWQGLRRVYFVDSFVQDPVELEVAQNALDGLLYPLIYGHFVDDRLAL	240	
Qy	241	EGGSAAWMLATFMPGEHTNESNRWTDVAVKTTWA----AESDDNRALLARWTRDWSARFAAA	296	

Db 241 OGGTAVAILTSFPEWHDETARWIDGVIKTMANGGEPEAAGNRALLSQWFSQWADRAQA 300
Qy 297 LAPVAARALQDAGRAALDEVEQFHARAARLGIAL 331
Db 301 LAPVAEMALGEGHAALGEVRAALDAAATKLGSL 335
RESULT 2
Q92NP6
ID Q92NP6 PRELIMINARY; PRT; 330 AA.
AC Q92NP6;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TEMBLrel. 10, Last annotation update)
DE PHENOL HYDROXYLASE COMPONENT.
GN APHL.
OS Comamonas testosteroni (Pseudomonas testosteroni).
OC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae; Comamonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TA441;
RX MEDLINE; 99018939.
RA ARAI H., AKAHIRA S., OHISHI T., MAEDA M., KUDO T.;
RT "Adaptation of Comamonas testosteroni TA441 to utilize phenol:
RT organization and regulation of the genes involved in phenol
RT degradation.";
RL Microbiology 144:2895-2903(1998).
DR EMBL; AB006479; BAA34170.1; -.
SQ SEQUENCE 330 AA; 37138 MW; 5566866E CRC32;

Query Match 63.9%; Score 1101.5; DB 2; Length 330;
Best Local Similarity 62.9%; Pred. No. 3.1e-93;
Matches 207; Conservative 43; Mismatches 78; Indels 1; Gaps 1;

Qy 1 MTIELKTVDIKPLRHTFAHVAQNIIGDKTATRYOEGMGAQPOENFHYRPTWDPDYIFD 60
Db 1 MNIELSAREIQPLRHTFAHVA--YTGDKPASRYLEATLGVPQTHFHYRPLWEPFELED 59
Qy 61 PSRAIRMANWYALKDPQFYASWATTRARQODAMESNFVESRRMIGLMDDDVAARA 120
Db 60 TGTAIRADWNLRDPROYYATWTTTRAKQOEAMANTQFVESRNLAKMSDALRAKA 119
Qy 121 LDVLVPLRHAAGANMNAQICGCTGTAPAMFHAMDNGLGVAOYLTRLALAMAPDV 180
Db 120 CGVLPURHVWANGNMNNSQICSGYCTPTAPAMFHAMDNGLGVAOYLTRLGLVMDPEGV 179
Qy 181 LEAKATWTRDAQAQPLRRYVEDTLVVDPVLEFIAQNLALDGLLYPLVYDRFVDERIAL 240
Db 180 LEAGKNDWQHAPWQPLRHLVLEDLLVVQDPLELFVAQNMVLDGLLYPLVYNHFVDERVT 239
Qy 241 EGGSAVAMLTAFMPWHTESNRWIDAVVKTMAESDDNRALLARWTRDWSARAALAPV 300
Db 240 OGGTAVAMLTAFMPWHTETARWIDAVVKTMAESDDNRALLARWTRDWSARAALAPV 299
Qy 301 AARALQDAGRAALDEVEQFHARAARLG 329
Db 300 AELALGERGAALSEVRQSLDIRAKAGL 328

RESULT 3
Q52570
ID Q52570 PRELIMINARY; PRT; 336 AA.
AC Q52570;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TEMBLrel. 08, Last annotation update)
DE ALPHA SUBUNIT-TERMINAL OXGENASE COMPONENT.
GN TMB.
OS Pseudomonas sp.
OC Bacteria; Proteobacteria.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-JS150;
RX MEDLINE; 96035667.
RA JOHNSON G.R., OLSEN R.H.;
RT "Nucleotide sequence analysis of genes encoding a toluene/benzene-2-
RT monooxygenase from Pseudomonas sp. strain JS150.";
RL Appl. Environ. Microbiol. 61:3336-3346(1995).
DR EMBL; L40033; AAA8457.1; -.
SQ SEQUENCE 336 AA; 37042 MW; E4D0F654 CRC32;

Query Match 59.3%; Score 1021.5; DB 2; Length 336;
Best Local Similarity 60.7%; Pred. No. 1.2e-76;
Matches 204; Conservative 40; Mismatches 87; Indels 5; Gaps 2;

Qy 1 MTIELKTVDIKPLRHTFAHVAQNIIGDKTATRYOEGMGAQPOENFHYRPTWDPDYIFD 60
Db 1 MNIDLQAREITPLRQTFHARLIGEGKAATRYQEAATYARSPWRTFTIARRGTPTSCTT 60
Qy 61 PSRAIRMANWYALKDPQFYASWATTRARQODAMESNFVESRRMIGLMDDDVAARA 120
Db 61 RAAAAIRLADCEALKDPQFYATWTTTRARQOEAVANFOFVESRGLASMGALDRRA 120
Qy 121 LDVLVPLRHAAGANMNAQICGCTGTAPAMFHAMDNGLGVAOYLTRLALAMAPDV 180
Db 121 LQVLLPLRHHVWANGNMNNAVICAYGCTGTAPAMFHAMDDLGIAQLITRLALTDEPAV 180
Qy 181 LEAKATWTRDAQAQPLRRYVEDTLVVDPVLEFIAQNLALDGLLYPLVYDRFVDERIAL 240
Db 181 LEAGKLAWEEDPRQGLRYVEDSFVVDPVLEFVAQNLALDGLLYPLIYGHFVDDHLAL 240
Qy 241 EGGSAVAMLTAFMPWHTESNRWIDAVVKTMAESDDNRALLARWTRDWSARAALAPV 296
Db 241 OGGTAVAILTSFPEWHDETARWIDAVIKTMANGGEPEAAGNRALLSQWFSQWADRAQA 300
Qy 297 LAPVAARALQDAGRAALDEVE-QFHARAARLGIAL 331
Db 301 LAPVAEMALGEGHAALGEVRAALDARATKLGSL 336

RESULT 4
O84959
ID O84959 PRELIMINARY; PRT; 331 AA.
AC O84959;
DT 01-NOV-1998 (TEMBLrel. 08, Created)
DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
DE PHENOL HYDROXYLASE COMPONENT.
GN POXB.
OS Ralstonia sp. E2.
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Ralstonia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-E2;
RX MEDLINE; 98361023.
RA HINO S., WATANABE K., TAKAHASHI N.;
RT "Phenol hydroxylase cloned from Ralstonia eutropha strain E2 exhibits
RT novel kinetic properties.";
RL Microbiology 144:0-0(0).
DR EMBL; AF026065; AAC32453.1; -.
SQ SEQUENCE 331 AA; 37592 MW; F273FAA0 CRC32;

Query Match 48.3%; Score 832; DB 2; Length 331;
Best Local Similarity 50.0%; Pred. No. 5e-61;
Matches 166; Conservative 49; Mismatches 113; Indels 4; Gaps 3;

Qy 1 MTIELKTVDIKPLRHTFAHVAQNIIGDKTATRYOEGMGAQPOENFHYRPTWDPDYIFD 60
Db 1 MQVDIKTQOIQPLRQTYGHVARRF-GDKPASRYQEATYDVQSEVNFHYRPTWAPQFELYD 59
Qy 61 PSRAIRMANWYALKDPQFYASWATTRARQODAMESNFVESRRMIGLMDDDVAARA 120

```
Db 60 KRRTAEMADWYALKDPQYYGAYVCTGRQOEAAEFKREKGLLQALPLEWERL 119
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 LDVLPPLRHAAGANNNAQICALGCTGVTFTAPAMFHAMDNGLGVAQYLTRLALAM--AEP 178
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 120 TDGLLLPLRHVEAANNMFYCADYGGTGTACTTCAMDRLGTAQYLSRIGLLDNTG 179
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 179 DVLEAAKATWTRDAAWQPLRRYVEDTLVADPVLELFTAQNALDGLLYPLVYDRFVDERI 238
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 180 AALERARAWLESEAWQPLRRYVEHSEFVIEDNFTQFTQNALDGLLYPLVY-CHADAAL 238
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 239 ALGGSAVAMLTAFMPPEWHTESNRWDVAVKTTMAAESDDNRALLARTRDWSARAEEALA 298
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 239 VRACGTGLAVVTEFMDNRDRHVRWVDVAVQVTAASEANRALLSRWGEARQAQAEALR 298
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 299 PVAARALODAGRAALDEVROQFHARAARLGI 330
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 299 PVAAILLGGEGEQAIACLEQFARLAKLGA 330
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
RESULT 5
Q52171 PRELIMINARY; PRT; 331 AA.
ID Q52171
AC Q52171;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE PHENOL HYDROXYLASE P1 PROTEIN (EC 1.14.13.7)
DE (PHENOL 2-MONOXYGENASE P1 COMPONENT).
GN PHLL.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P35X / NCBI 9869;
RX MEDLINE; 95129877.
RA NG L.C., SHINGLER V., SZE C.C., POH C.L.;
RT "Cloning and sequences of the first eight genes of the chromosomally
RT encoded (methyl) phenol degradation pathway from Pseudomonas putida
RT P35X."
RL Gene 151:29-36(1994).
CC -1- FUNCTION: CATABOLIZES PHENOL, AND SOME OF ITS METHYLATED
CC DERIVATIVES.
CC -1- CATALYTIC ACTIVITY: PHENOL + NADPH + O(2) = CATECHOL + NADP(+) +
CC H(2O).
CC -1- COFACTOR: FAD FLAVOPROTEIN, AND REQUIRES FE(+2) FOR ACTIVITY (BY
CC SIMILARITY).
CC -1- PATHWAY: FIRST STEP OF PHENOL BIODEGRADATION.
CC -1- SUBUNIT: THE MULTICOMPONENT ENZYME PHENOL HYDROXYLASE IS FORMED BY
CC P0, P1, P2, P3, P4 AND P5 POLYPEPTIDES.
CC EMBL; X79063; CAA55661.1; -.
KW Aromatic hydrocarbons catabolism; Oxidoreductase; Monooxygenase;
KW Flavoprotein; FAD; Iron.
SQ SEQUENCE 331 AA; 38178 MW; FD7E5B1A1 CRC32;
```

Query Match 45.2%; Score 778.5; DB 2; Length 331;
Best Local Similarity 46.8%; Pred. No. 1.3e-56;
Matches 155; Conservative 54; Mismatches 115; Indels 7; Gaps 4;

```
QY 1 MTELTQVDIKPLRHFAHVAQNIIGDKTATRYQEGMGAQPOENFYRTWDPDYEIFD 60
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSVEIKTNVDPTRQTYGNLQRRF-GDKPASRYQEAASYDIEAVTNFYRPLWDPQHELHD 59
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 PSRSATRWANNYALKDPQRYFYASWATTRARQODAMESNFEEVSRMRIGLMDRDVAA 120
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 60 PTTAINTDWHKVTDPRQYFYAYVQTRARMQEAETHAYGFCREKRELLSRPAELQAKL 119
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 LDVLPPLRHAAGANNNAQICALGCTGVTFTAPAMFHAMDNGLGVAQYLTRLALAM--AEP 178
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 120 LRLVPLRHAELGANNNNSSIAGDSIAATVTQHIYQAMDRLMGQYLSRIGLLDGGTG 179
```

```
QY 179 DVLEAAKATWTRDAAWQPLRRYVEDTLVADPVLELFTAQNALDGLLYPLVYDRFVDERI 238
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 180 EALDQAKAYWLDPIWQGLRYYVEDSFVIRDFELGLAQNLVLDGLLOPLMYQRF-DOWL 238
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 239 ALGGSAVAMLTAFMPPEWHTESNRWDVAVKTTMAAESDDNRALLARTRDWSARAEEALA 298
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 239 TENGSDVAMLTETFRMDRWYGESTRWDAFKTVLAENDANREQVQAWLEVPEPRAYEALL 298
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 299 PVAARALODAGRAALDEVROQFHARAARLGI 329
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 299 PLABEA---TGIALDEVRSFAFATRLQKIGL 326
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
RESULT 6
Q52162 PRELIMINARY; PRT; 342 AA.
ID Q52162
AC Q52162;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE PHENOL HYDROXYLASE P1 PROTEIN (EC 1.14.13.7)
DE (PHENOL 2-MONOXYGENASE P1 COMPONENT).
GN PHLB.
OS Pseudomonas putida.
OG Plasmid pPGH1.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE; 95272534.
RA HERRMANN H., MUELLER C., SCHMIDT I., MAHNKE J., PETRUSCHKA L.,
RA HAHNKE K.;
RT "Localization and organization of phenol degradation genes of
RT Pseudomonas putida strain H."
RL Mol. Gen. Genet. 247:240-246(1995).
CC -1- FUNCTION: CATABOLIZES PHENOL, AND SOME OF ITS METHYLATED
CC DERIVATIVES.
CC -1- CATALYTIC ACTIVITY: PHENOL + NADPH + O(2) = CATECHOL + NADP(+) +
CC H(2O).
CC -1- COFACTOR: FAD FLAVOPROTEIN, AND REQUIRES FE(+2) FOR ACTIVITY (BY
CC SIMILARITY).
CC -1- PATHWAY: FIRST STEP OF PHENOL BIODEGRADATION.
CC -1- SUBUNIT: THE MULTICOMPONENT ENZYME PHENOL HYDROXYLASE IS FORMED BY
CC P0, P1, P2, P3, P4 AND P5 POLYPEPTIDES.
CC EMBL; X80765; CAA36741.1; -.
KW Aromatic hydrocarbons catabolism; Oxidoreductase; Monooxygenase;
KW Flavoprotein; FAD; Iron; Plasmid.
SQ SEQUENCE 342 AA; 39370 MW; 430445B9 CRC32;
```

Query Match 44.7%; Score 770.5; DB 2; Length 342;
Best Local Similarity 46.5%; Pred. No. 6.1e-56;
Matches 154; Conservative 54; Mismatches 116; Indels 7; Gaps 4;

```
QY 1 MTELTQVDIKPLRHFAHVAQNIIGDKTATRYQEGMGAQPOENFYRTWDPDYEIFD 60
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 12 MSVEIKTNVDPTRQTYGNLQRRF-GDKPASRYQEAASYDIEAVTNFYRPLWDPQHELHD 70
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 PSRSATRWANNYALKDPQRYFYASWATTRARQODAMESNFEEVSRMRIGLMDRDVAA 120
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 71 PTTAINTDWHKVTDPRQYFYAYVQTRARMQEAETHAYGFCREKRELLSRPAELQAKL 130
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 LDVLPPLRHAAGANNNAQICALGCTGVTFTAPAMFHAMDNGLGVAQYLTRLALAM--AEP 178
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 131 LRLVPLRHAELGANNNNSSIAGDSIAATVTQHIYQAMDRLMGQYLSRIGLLDGGTG 190
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 179 DVLEAAKATWTRDAAWQPLRRYVEDTLVADPVLELFTAQNALDGLLYPLVYDRFVDERI 238
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 191 EALDQAKAYWLDPIWQGLRYYVEDSFVIRDFELGLAQNLVLDGLLOPLMYQRF-DOWL 249
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 239 ALGGSAVAMLTAFMPPEWHTESNRWDVAVKTTMAAESDDNRALLARTRDWSARAEEALA 298
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 250 TENGSDVAMLTFFRDMYGESTRWVDMFKTVLAENDANREQVQAWLEWPEPRAYEALL 309
QY 299 PVAARALQDAGRAALDEVREQPHARAALGI 329
Db 310 PLAEER- -TGIAALDEVRSFAFATRLQKIGL 337
RESULT 7
Q43979
ID Q43979 PRELIMINARY; PRT; 333 AA.
AC Q43979;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE PHENOLHYDROXYLASE COMPONENT.
OS Acinetobacter calcoaceticus.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Moraxellaceae; Acinetobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCIB8250;
RX MEDLINE; 96154937.
RA EHRT S., SCHIRMER F., HILLEN W.;
RT "Genetic organization, nucleotide sequence and regulation of
RT expression of genes encoding phenol hydroxylase and catechol 1,2-
RT dioxygenase in Acinetobacter calcoaceticus NCIB8250.";
RL Mol. Microbiol. 18:13-20(1995).
DR EMBL; 236909; CA85381.1; -.
SQ SEQUENCE 333 AA; 38382 MW; 83055EC9 CRC32;

Query Match 37.2%; Score 640.5; DB 2; Length 333;
Best Local Similarity 39.4%; Pred. No. 3e-45;
Matches 127; Conservative 70; Mismatches 120; Indels 5; Gaps 4;

QY 1 MTIELKTVDIKPLRHTFAHVAQNIIGDKTATRYQEGMGAQPOENFHYRPTWDPDYEIFD 60
Db 1 MTLEIKTSNLEPIRQTYAYIERRFGA-KPATRYQEVSPDIOASTNHFYRPLWKPKDKTLND 59
QY 61 PPSAIRMANWYALKDPQFYFASWATTRARQODAMESNFEFESRRMIGLMRDVVAARA 120
Db 60 KHTHALQMDWYAFKDPQFYGYVQHRARLQDTAESHYAFFEKRQLVNNLSDEVKOKI 119
QY 121 LDVLPVLRHAAGANNNAQICALCYGVFTAPAMFAMDMNLGVAQYLRALALAM--AEP 178
Db 120 IOCLLPFRYVEGTANLHMMSGAYGYGVITQACIFAAMDRLGMAQYISRGILIDGNTG 179
QY 179 DVLEAAKATWTRDAWQPLRYVEDTLVADPVELFTIAQNALDGLLYPLVYDRFVDRI 238
Db 180 ESLOQAKHAWLNDETQWPLRLCEQSLTEQDFWFKLYILQNLIDSLMLQELVFGQ-LDEWL 238
QY 239 ALEGGSAVAMLTAFMPEWHSTESNRWIDAVVKTMAESDDNRALLARWTRDSARAEALA 298
Db 239 VENGGRDIAITEFMKOCCLDLAKWSDSVLTAISESDNKTLIQSITELLPOVKQAFS 298
QY 299 PVAARALQDAG-RAALDEVREQ 319
Db 299 AWAQALTDSGIDSLNKAISER 320
RESULT 8
O32429
ID O32429 PRELIMINARY; PRT; 333 AA.
AC O32429;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE DNS OXYGENASE COMPONENT.
OS Acinetobacter sp.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Moraxellaceae; Acinetobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCIB8250;
RX MEDLINE; 96154937.
RA BERTONI G., MARTINO M., GALLI E., BARBIERI P.;
RT "Analysis of the gene cluster encoding toluene/o-xylene monooxygenase
RT from Pseudomonas stutzeri OX1.";
RL Appl. Environ. Microbiol. 64:3626-3632(1998).
DR EMBL; AJ005663; CAA06658.1; -.
SQ SEQUENCE 330 AA; 38334 MW; 4EFFIAD8 CRC32;

Query Match 12.9%; Score 222.5; DB 2; Length 330;
Best Local Similarity 25.1%; Pred. No. 8.2e-11;
Matches 83; Conservative 52; Mismatches 147; Indels 49; Gaps 14;

QY 10 IKPLRHRTFAHVAQNIIGDKTATRYQEGMGAQPOENFHY-----RP-----TWD 53
Db 9 LKPLK-TWSHIA---GNRRRSEYE-----VSTNLHYFTDNPERPWELDSNLPMTQWY 58
```

```
RP SEQUENCE FROM N.A.
RC STRAIN-20B;
RX MEDLINE; 98005684.
RA HORINOUCHI M., KASUGA K., NOJIRI H., YAMANE H., OMORI T.;
RT "Cloning and characterization of genes encoding an enzyme which
RT oxidizes dimethyl sulfide in Acinetobacter sp. strain 20B.";
RL FEMS Microbiol. Lett. 155:99-105(1997).
DR EMBL; D85083; BAA23331.1; -.
SQ SEQUENCE 333 AA; 38393 MW; BCL06309 CRC32;

Query Match 37.2%; Score 640.5; DB 2; Length 333;
Best Local Similarity 38.7%; Pred. No. 3e-45;
Matches 129; Conservative 69; Mismatches 128; Indels 7; Gaps 4;

QY 1 MTIELKTVDIKPLRHTFAHVAQNIIGDKTATRYQEGMGAQPOENFHYRPTWDPDYEIFD 60
Db 1 MTLEIKTSNLEPIRQTYAYIERRFGA-KPATRYQEVSPDIOASTNHFYRPLWKPKDKTLND 59
QY 61 PPSAIRMANWYALKDPQFYFASWATTRARQODAMESNFEFESRRMIGLMRDVVAARA 120
Db 60 KHTHALQMDWYAFKDPQFYGYVQHRARLQDTAESHYAFFEKRQLVNNLSDEVKOKI 119
QY 121 LDVLPVLRHAAGANNNAQICALCYGVFTAPAMFAMDMNLGVAQYLRALALAM--AEP 178
Db 120 IOCLLPFRYVEGTANLHMMSGAYGYGVITQACIFAAMDRLGMAQYISRGILIDGNTG 179
QY 179 DVLEAAKATWTRDAWQPLRYVEDTLVADPVELFTIAQNALDGLLYPLVYDRFVDRI 238
Db 180 ESLOQAKHAWLNDETQWPLRLCEQSLTEQDFWFKLYILQNLIDSLMLQELVFGQ-LDEWL 238
QY 239 ALEGGSAVAMLTAFMPEWHSTESNRWIDAVVKTMAESDDNRALLARWTRDSARAEALA 298
Db 239 VENGGRDIAITEFMKOCCLDLAKWSDSVLTAISESDNKTLIQSITELLPOVKQAFS 298
QY 299 PVAARALQDAGRAALDEVREQPHARAALGI 331
Db 299 AWAQALTDSG---IDSLNKAISERSKKAGTIL 328
RESULT 9
O87802
ID O87802 PRELIMINARY; PRT; 330 AA.
AC O87802;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE TOLUENE, O-XYLENE MONOOXYGENASE OXYGENASE SUBUNIT.
GN TOUE.
OS Pseudomonas stutzeri (Pseudomonas perfectomarina).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OX1;
RX MEDLINE; 98432776.
RA BERTONI G., MARTINO M., GALLI E., BARBIERI P.;
RT "Analysis of the gene cluster encoding toluene/o-xylene monooxygenase
RT from Pseudomonas stutzeri OX1.";
RL Appl. Environ. Microbiol. 64:3626-3632(1998).
DR EMBL; AJ005663; CAA06658.1; -.
SQ SEQUENCE 330 AA; 38334 MW; 4EFFIAD8 CRC32;

Query Match 12.9%; Score 222.5; DB 2; Length 330;
Best Local Similarity 25.1%; Pred. No. 8.2e-11;
Matches 83; Conservative 52; Mismatches 147; Indels 49; Gaps 14;

QY 10 IKPLRHRTFAHVAQNIIGDKTATRYQEGMGAQPOENFHY-----RP-----TWD 53
Db 9 LKPLK-TWSHIA---GNRRRSEYE-----VSTNLHYFTDNPERPWELDSNLPMTQWY 58
```


QY 54 PDYEFDPSSRAIRMANWYALKDPQFYASWATTRARQDAMESNFEFESRRMIGLR 113
 Db 59 KKY-CFD---SPLKDDNNAFRDPQLVYRTYNLLQDQESYVQGLFDQLNDRGHQMLT 114
 QY 114 DVVAARALDVLVLPURHAAGANMNAQICALGYCTVTAPAMFHAMDNGLGVAQYLRLA- 172
 Db 115 REWVETLAREYTPARYLFHALQMSVYIHQIAPASTITNCATYETADHL---RWLTHWAY 171
 QY 173 ----LMAAEPDV--LEAAKATWTDRDAWQPLRRYVEDTLVADPVELFIAONLALDGLLY 226
 Db 172 RRELANCYFDVGGKRRERDWEWNPDAWQGFRELIEKALTAMDGEFTALNL-----VTK 227
 QY 227 PLVYDFVDERIAL---EGGSAVAMLTAFMPWHTESNRWIDAYVTKMAESDDNRALLA 283
 Db 228 PAVEEALLQGLSLAQSEGDTLLGLLAQAOKRDAERHRRWSSALVK-MALEKEGNREVIQ 286
 QY 284 RWTRDWSARAEEALAPVAARALQDAGRAALD 314
 Db 287 KWVAKWEPLADKAIEAYCS-ALPDGENAIVE 316

RESULT 10

Q51943

ID Q51943 PRELIMINARY; PRT; 329 AA.

AC Q51943

DT 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)

DE TOLUENE-3-MONOOXYGENASE OXYGENASE SUBUNIT.

GN TBUA2.

OS Burkholderia pickettii (Pseudomonas pickettii).

OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;

OC Ralstonia.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-PRO1.

RX MEDLINE; 95172404.

RA BYRNE A.M., KUKOR J.J., OLSEN R.H.;

RT "Sequence analysis of the gene cluster encoding toluene-3-

RT monooxygenase from Pseudomonas pickettii PKO1.";

RL Gene 154:65-70(1995).

DR EMBL; U04052; AAB09622.1; -.

KW Monooxygenase.

SQ SEQUENCE 329 AA; 37525 MW; C61B3727 CRC32;

Query Match

Best Local Similarity 11.2%; Score 193; DB 2; Length 329;

Matches 88; Conservative 55; Mismatches 132; Indels 82; Gaps 18;

QY 10 IKPLRHTFAHVAQNIIGDKTATRYOEGMGAQPOE-----NFHYRP-----TWDPDY 56
 Db 8 LKPLK-TWSHLA-----ARR-----RKPEYEIVSTNLHYTTDNDPAPFELDPNF 51
 QY 57 EI----FDPSSRA--IRMANWYALKDPQFYASWATTRARQDAMESNFEFESRRMIGL 111
 Db 52 EMAQWFKRNRNASPLTHPDNNAFDPDELVYRTYNML---QDQET-----YFVGL 99
 QY 112 MRDDVAARALDVLV-----PLRHAAGANMNAQICALGYCTVTAPAMFHAMDNGLGVAQYLRLA- 158
 Db 100 L-DQFSRGHDAMLERTWAGTFLARLYTPVRYLFHTLQMSAYLTQAPASTISNCAAYQT 158
 QY 159 MDNLGVAQYLTRLALAAEPDVLEA-----AKATWTRDAWQPLRRYVEDTLVADPV 211
 Db 159 ADSL---RWLTHAYRTKLSQTSFSDLGFGTDERRYWEQDPAMQGWKRLVEHALVANDWA 215
 QY 212 ELFTAONLALDGLLYPLVYD---RFVDERIALEGGSAMLTAFMPWHTESNRWIDAVV 268
 Db 216 ECFVALSL-----VVRPAVEAVLRSIGEARHNGDTLLGLLTDQAQADQHRHWAGALV 271
 QY 269 KTMMAESDDNRALLARWTRWSARAEEALAPVAARALQDAGRA---ALDEVREQFHA 322
 Db 272 R-MALEQPGNREVITGWLKWEPLADEAIVAYCS-ALPEAPAAQARATAAVREFRHS 326

RESULT 11

O69182

ID O69182 PRELIMINARY; PRT; 332 AA.

AC O69182

DT 01-AUG-1998 (Tremblrel. 07, Created)

DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)

DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)

DE PUTATIVE HYDROXYLASE COMPONENT.

GN PHLO.

OS Alcaligenes eutrophus.

OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;

OC Ralstonia.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-JMP134;

RA HARKER A.R., AYOUBI P.J.;

RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF065891; AAC77384.1; -.

SQ SEQUENCE 332 AA; 37511 MW; 871144CB CRC32;

Query Match

Best Local Similarity 11.1%; Score 190.5; DB 2; Length 332;

Matches 82; Conservative 53; Mismatches 145; Indels 47; Gaps 13;

QY 10 IKPLRHTFAHVAQNIIGDKTATRYOEGMGAQPOE-----NFHYRP-----TWDPDY 56
 Db 8 LKPLK-TWSHLA-----ARR-----RKPEYEIVSTNLHYTTDNDPAPFELDPNF 51
 QY 57 EI----FDPSSRA--IRMANWYALKDPQFYASWATTRARQDAMESNFEFESRRMIGL 111
 Db 52 EMAQWFKRNRNASPLKHDWNAFDPDELVYRTYNMLQDQETVSGLLDQFSRGHDSM 111
 QY 112 MRDDVAARALDVLVPLRHAAGANMNAQICALGYCTVTAPAMFHAMDNGLGVAQYLRL 171
 Db 112 LEHSWAGTFLARLYTPARYLFHALQMSAYLTQMAPASTISNCAAYQTADSL---RWLTH 168
 QY 172 A-----LMAAEPDVLEAA--KATWTRDAWQPLRRYVEDTLVADPVELFIAONLALDGL 224
 Db 169 AYRTRELSTQFGDVGFGTDERRYWEQDPAMQGWKRLVEHALVANDWAECFVAFSLVRPA 228
 QY 225 LYPLVYDFVDERIALEGGSAMLTAFMPWHTESNRWIDAVVTKMAESDDNRALLAR 284
 Db 229 MEEAVL-RGUEAARHNGDTLLGLLTDQAQADQHRHWAGALVR-MALETTPGNRDLVAG 286
 QY 285 WTRDWSARAEEALAPVAARALQDAGRA 311
 Db 287 SIARWAPLADDAISAYCA-ALPDAPNA 312

RESULT 12

Q53025

ID Q53025 PRELIMINARY; PRT; 343 AA.

AC Q53025

DT 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DT 01-JAN-1999 (Tremblrel. 09, Last annotation update)

DE EPOXIDASE SUBUNIT.

GN AMOA.

OS Nocardia corallina.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Nocardiaceae; Nocardia.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-B-276;

RA SAEKI H., FURUHASHI K.;

RT "Cloning and characterization of a Nocardia corallina B-276 gene

RT cluster encoding alkene monooxygenase.";

RL J. Ferment. Bioeng. 78:339-406(1994).

DR EMBL; D37875; BAA07112.1; -.

SQ SEQUENCE 343 AA; 38462 MW; 9195C333 CRC32;

	Query Match	10.4%; Score 179.5; DB 2; Length 343;
	Best Local Similarity	22.9%; Pred. No. 3e-07;
	Matches	77; Conservative 55; Mismatches 145; Indels 59; Gaps 14;
QY	31 TRYQEGMGAQ--PQENPHY-----RPTWDPDYEIFDPSAIRMANWYALKDRPQ	79
Dd	30 TEXELYTVGQOSTPDWEHLVDMPLRFDDGRAPWEE-----SSAVTSEWSAYRDPHQ	82
QY	80 FYASAWATRARQQDAMESNEFEVESRMICGLMRDDVAARALDLVLPL-----RHAA	131
Dd	83 LWQRPYSTCNDQOAL-----ARLVPVLTNGSAA-----ITPIWSOKILARSAA	128
QY	132 W-----GANMNAQICALGYGTFTAPAMFTAMDNLGVAQYLTLALAMAEPDVLEA-A	184
Dd	129 WPFEYGLFLSLAYAVROAMSDTQFSVVFOAVDRMRLQLDIVHLDHLQESPEFSDAGA	188
QY	185 KATWTROMAQOPRRYVEDTLVADPVLFETIAQNALDGLLYPLVYDFVDERTALEG-G	243
Dd	189 REANMSDSTLVPIREVIERIAASODWBELVAGTLVFPELVGHAKAELEFSRRAPMFGDG	248
QY	244 SAVAMLTAFMPE--WHRESNWIDAVVKTMARE---SDNRALLARWTWRDSARAEEA--	296
Dd	249 TTPAVLASALDSGRHLES-----VOALRVLYQCDDPVHGDNQNAVRRWTEEWOPRCKAAAQ	305
QY	297 -LAPVAARALQDAGRALDEVREQFHARAARGIALG	331
Dd	306 SFLTFFDCGIDAKESANALSALANQRAAVEGAGI	341
RESULT	13	
ID	Q9ZET3	PRELIMINARY; PRT: 341 AA.
AC	Q9ZET3;	
DT	01-MAY-1999 (TrEMBLrel. 10, Created)	
DT	01-MAY-1999 (TrEMBLrel. 10, Last sequence update)	
DT	01-MAY-1999 (TrEMBLrel. 10, Last annotation update)	
DE	OXYGENASE BETA SUBUNIT.	
GN	XAMOE.	
OS	Xanthobacter sp. py2.	
OC	Bacteria; Proteobacteria; alpha subdivision; Ancylobacter group;	
CC	Xanthobacter.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-PY2;	
RA	ZHOU N.Y., CHAN KWONG C.K., LEAK D.J.;	
RT	"The alkene monooxygenase from Xanthobacter Py2 is closely related to aromatic monooxygenase and catalyzes aromatic monohydroxylation of benzene, toluene and phenol."	
RL	Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.	
DR	EMBL: AJ012090; CAA09915.1; ..	
SQ	SEQUENCE 341 AA; 38223 MW; 8775471D CRC32;	
Query Match	9.7%; Score 167.5; DB 2; Length 341;	
Best Local Similarity	26.1%; Pred. No. 2.9e-06;	
Matches	74; Conservative 49; Mismatches 124; Indels 37; Gaps 14;	
QY	65 AIRMANWYALKDRPQFYASWATTRARQQDAMESNEFEVESRRMI---- <td>120</td>	120
Dd	65 ALKVDSWDLFRDPRTTY----DTYVMQDDQET---YVDNLLLSYTGEGRYDELSSRS	117
QY	121 LDV-----LVPLRHAACANNNAQICALGYGTFTAPAFHAMDNLGVAQ----VLTRLAL	173
Dd	118 LDDLASGLTPTRYLGHGLQMLAAYIQOLAPSAYVGVNCVAFQTSDALRRVRVAYRTR-QL	176
QY	174 AMAEP--DVLEAAKATWTRDAWAQPLRRYVEDTLVADPVLEFETIAQNALDGLLYPLVID	231
Dd	177 ADAHPARGFGSGDRANVEKSPDKQPIKAIEELLVTENDKALAGTFNVKPKPIDELFLN	236
QY	232 RFVDERTALEGGSAVAMLTAFMPPEWHITESN---RWIDAVVKVTMAAESDDRALLARWTRD	288

RX MEDLINE; 97288080.
RA MCDONALD I.R., UCHIYAMA H., KAMBE S., YAGI O., MURRELL J.C.;
RT "The soluble methane monooxygenase gene cluster of the
RL trichloroethylene-degrading methanotroph Methylocystis sp. strain M.";
DR Appl. Environ. Microbiol. 63:1898-1904(1997).
RL EMBL; U81594; AAC45290.1;
KW Monooxygenase.
SQ SEQUENCE 395 AA; 45170 MW; E4507FBD CRC32;

Query Match 7.1%; Score 121.5; DB 2; Length 395;
Best Local Similarity 20.8%; Pred. No. 0.022;
Matches 68; Conservative 50; Mismatches 164; Indels 45; Gaps 12;
QY 28 KTATRYQEGMGAQPO-----ENFH-YRPTWDPDYEIFDPSPSAIRMANWY 72
DB 47 KRUSEYQSLSCYAPNPDPNIAGGLDWDGWTQKFGHGRPSGNE-----STELRTTDWY 99
QY 73 ALKDPQFYASWA---TTRARQDAMESNPEFVESRRMIG-LMRDDVAARALDVLPLR 128
DB 100 RHRDPARRWHAPYVKDSEEARYTQRFLLAAYSSEGSIRTIDPYWRDEILNKYYGALLYNE 159
QY 129 HAANGANNNAQICALGYGTFTAPAMFAMDNLGVAQ--YLTRLALAMAEPDV---LEA 183
DB 160 YGLFNAHSSVGRDC---LSDTIRQSATFAGLDKVDNAQMIQMERLFIKLVPGFDASTDV 216
QY 184 AKATWTRDAAWQPLRRYVEDTLV-VADPVELFIAQNALDGLLYPLVYDRFVDERIALEG 242
DB 217 PKIWTSDPIYAGARGAVEEIIWQIQDWNELIHWAGHAVYDATFGQFARREFQRLATVYG 276
QY 243 GSAVAMLTAFMPEWHTESNRWIDAVVKTMAESDD---NRALLARWTRDWSARAAL- 297
DB 277 DTLTPFTAQSOITYFTQTRGAIDDLFYCYCLANDPEFGAHNRTFLNATHEYLARSVTALK 336
QY 298 --APVAARALQDAG---RAALDEVREQ 319
DB 337 DFVGIIYAKVEKVAGATDRAGVSEALQR 363

Search completed: September 26, 2000, 20:28:52
Job time: 1011 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:22:37 ; Search time 60.65 seconds
(without alignments)
22.495 Million cell updates/sec

Title: US-09-430-029-4

Perfect score: 451

Sequence: 1 MSNVEIAFOANEDSRPIVDA.....ITLSGHIDDDDEFTLSWSH 89

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues

Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6_COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	355	78.7	92	1	US-08-319-387-3
2	122	27.1	103	5	5171684-5
3	65	14.4	1162	2	US-08-663-566A-15
4	65	14.4	1162	2	US-08-023-610-15
5	65	14.4	1162	2	US-08-288-065A-15
6	65	14.4	1162	2	US-08-362-240A-15
7	65	14.4	1162	4	PCT-US95-10245-15
8	64	14.2	292	3	PCT-US85-190-3
9	62.5	13.9	406	5	5212296-6
10	62	13.7	322	2	US-08-505-218-6
11	62	13.7	897	1	US-07-960-389-2
12	61	13.5	256	3	US-08-306-769-89
13	61	13.5	256	4	PCT-US95-14442A-89
14	59.5	13.2	339	1	US-08-266-451B-22
15	59.5	13.2	339	2	US-08-748-725-22
16	59.5	13.2	464	1	US-08-475-894-4
17	59.5	13.2	464	1	US-08-484-710-4
18	59.5	13.2	464	2	US-08-484-709-4
19	59.5	13.2	553	1	US-08-475-894-2
20	59.5	13.2	553	1	US-08-484-710-2
21	59.5	13.2	553	2	US-08-484-709-2
22	58	12.9	117	1	US-08-499-215-3
23	58	12.9	682	2	US-08-436-900A-4
24	58	12.9	1477	3	US-08-492-459-10
25	58	12.9	1477	3	US-08-423-752-10
26	58	12.9	1477	3	US-08-945-994-3
27	55.5	12.3	381	2	US-08-387-942C-25
28	55.5	12.3	425	1	US-08-190-802A-48

29 55.5 12.3 425 3 US-08-105-454-10 Sequence 10, Appli
30 55.5 12.3 575 2 US-08-766-858A-5 Sequence 5, Appli
31 55.5 12.3 1403 2 US-08-387-942C-3 Sequence 3, Appli
32 55.5 12.3 2544 2 US-08-576-626A-32 Sequence 32, Appli
33 55 12.2 308 2 US-08-484-905-78 Sequence 78, Appli
34 55 12.2 308 3 US-08-481-985B-78 Sequence 78, Appli
35 55 12.2 484 1 US-08-597-236-2 Sequence 2, Appli
36 55 12.2 484 1 US-08-746-682A-2 Sequence 2, Appli
37 55 12.2 672 1 US-08-049-254-2 Sequence 2, Appli
38 55 12.2 672 1 US-08-472-934-2 Sequence 2, Appli
39 55 12.2 672 2 US-08-323-460A-2 Sequence 2, Appli
40 55 12.2 672 2 US-08-461-145C-2 Sequence 2, Appli
41 55 12.2 672 3 US-08-461-145C-2 Sequence 2, Appli
42 54 12.0 115 2 US-08-487-031-15 Sequence 15, Appli
43 54 12.0 853 3 US-08-451-946B-6 Sequence 6, Appli
44 54 12.0 853 3 US-08-446-938B-6 Sequence 6, Appli
45 54 12.0 853 3 US-08-446-939B-6 Sequence 6, Appli

ALIGNMENTS

RESULT 1

US-08-319-387-3

; Sequence 3, Application US/08319387

; Patent No. 5543317

; GENERAL INFORMATION:

; APPLICANT: Shields, Malcolm S.

; APPLICANT: Francesconi, Stephen C.

; TITLE OF INVENTION: Microbial Degradation of Trichloroethylene,

; TITLE OF INVENTION: Dichloroethylenes and Aromatic Pollutants

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Saliwanchik & Saliwanchik

; STREET: 2421 N.W. 41st Street, Suite A-1

; CITY: Gainesville

; STATE: FL

; COUNTRY: USA

; ZIP: 32606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA: US/08/319,387

; APPLICATION NUMBER: US/08/319,387

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/167,457

; FILING DATE: 15-DEC-1993

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/694,718

; FILING DATE: 02-MAY-1991

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Whitlock, Ted W.

; REGISTRATION NUMBER: 36,965

; REFERENCE/DOCKET NUMBER: UWF-2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 904-375-8100

; TELEFAX: 904-372-5800

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 92 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; US-08-319-387-3

Query Match 78.7%; Score 355; DB 1; Length 92;
Best Local Similarity 73.0%; Pred.No. 1.8e-39;
Matches 65; Conservative 14; Mismatches 10; Indels 0; Gaps

Qy	1 MSNVIAFOANEOSRPVDAIVADNPRAVVESPGMWKIDAPDLRTIRRETIETLTGRF 60
Dd	3 MSNVIAFOANEESPVEALIANPRAVATESPGMWKIDAPGLHTINROSIEDRIGMKF 62

QY	61	DLQQLQVNLITLSGHIDEDEDDFTLSWSH	89
		: : :	
Db	63	DLQQIHNLITLSGYIDEDEDDFTLSWKH	91

RESULT 2
5171684-5
; Patent No. 5171684

; APPLICANT: YEN, KWANG-MU; BLATT, LAWRENCE M.; KARL, MICHAEL R.
 ; TITLE OF INVENTION: BIOCONVERSIONS CATALYZED BY THE TOULUENE
 ; MONOOXYGENASE OF PSEUDOMONAS MENDOCINA KR-1

```

; NUMBER OF SEQUENCES: 41
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/590,374
; FILING DATE: 28-SEP-1999

```

FILING DATE: 28-SEP-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 177,631
 FILING DATE: 05-APR-1988

```

; SEQ ID NO: 5
; LENGTH: 103
5171684-5

```

Query Match 27.1%; Score 122; DB 5; Length 103;
Best Local Similarity 32.5%; Pred. NO. 7.4e-09;
Matches 27: Conservative 21; Mismatches 33; Indels 2; Gaps

QY 2 SNVFIAEQANESRPIVDAIVADNP-RAVVES-PGMVKIDAPDLTIRRETIEELTCTR 59

DB	12	NNVGPIIRAGDLVEPVIEIAEIDNPGKIEITVEDRRAYVRIAAEGELILTRKLTLEEQLGRP	71
QY	60	FDLQQLQVNLTLSGCHIDEEDDE	82

Db 72 FMQLEINLAFAGIQADEQ 94

RESULT 3
US-08-663-566A-15
; Sequence 15, Application US/08663566A

; PATENT NO. 5853733
 ;
 ; GENERAL INFORMATION:
 ;
 ; APPLICANT: Cochran, Mark D
 ;
 ; APPLICANT: Macdonald, Richard D

;	TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys
;	
;	TITLE OF INVENTION: and Uses Thereof
;	
;	NUMBER OF SEQUENCES: 56

;;
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: John P. White
;; STREET: 1185 Avenue of the Americas

CITY: NEW YORK
STATE: New York
COUNTRY: USA
ZIP: 10016

```

;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
;

```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/663,566A
FILING DATE: June 13, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: White, John P
REGISTRATION NUMBER: 28,678

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0526

```

; ID: 1
; NAME: 1162
; LENGTH: 1162
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1162 amino acids

```

```

;
; TYPE: amino acid
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein

```

US-08-663-566A-15

Query Match	14.48; S
Best Local Similarity	26.78; P
Matches	28; Conservative 11.

QY 4 VFIAFQANEDSRPIVDAIVA-----
||| | | | | |

Db 934 VFIHFSYTPDSFVNVTAI VGFVCVKP
QY 49 RETIEEL-TGTRFDLQQLVNLITL

Db 994 ATAGDIVTLTSCQANYVSVNKTVI

RESULT 4
US-08-023-610-15
; Sequence 15, Application US/08023

Patent No. 5928648
GENERAL INFORMATION:
APPLICANT: Cochran Ph.D, Mark

```
; APPLICANT: Macdonald Ph.D., R
; TITLE OF INVENTION: Recombinaa
; TITLE OF INVENTION: and Uses
; NUMBER OF SEQUENCES: 47
```

NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 30 Rockefeller Plaza

CITY: New York
STATE: New York
COUNTRY: USA

```

; ZIP: 10112
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS
; SOFTWARE: PatentIn Release
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/00
FILING DATE: February 26, 19
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: White Esq, John P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)977-9550

TELEFAX: (212)664-0525
TELEX: 422523
INFORMATION FOR SEQ ID NO: 15:

```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1162 amino acids
; TYPE: amino acid
;

```

; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-023-610-15

Query Match	14.4%; S
Best Local Similarity	26.7%; P

Matches 28; Conservative 11;
QY 4 VFIAFOANEDSRPIVDAIVA-----

[illegible]

Db 934 VFHFSTPDSFVNVTAVGFCVKPANASQYAIVPANGRGIFIQVNGSYIITARDMYMPR 993
Qy 49 RETIEEL-TGTRFDLQQLVNLTLSGHIDEDD-----DEFTLSWS 88
Db 994 AITAGDIVTLTSCQANYVSVNKTIVITTFVDNDDFDNDELSKWN 1038

RESULT 5
US-08-288-065A-15
; Sequence 15, Application US/08288065A
; Patent No. 5961982
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D
; APPLICANT: Macdonald, Richard D
; TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys S-
; TITLE OF INVENTION: HVT-050 and Uses Thereof
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/288.065A
; FILING DATE: Aug-09-94
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)278-0400
; TELEFAX: (212)391-0526
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1162 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-288-065A-15

Query Match 14.4%; Score 65; DB 2; Length 1162;
Best Local Similarity 26.7%; Pred. No. 6.7;
Matches 28; Conservative 11; Mismatches 46; Indels 20; Gaps 3;

Qy 4 VFIAFQANEDSRPIYDAIVA-----DNPRVVVSPGMVKIDAPDRLTIR 48
Db 934 VFHFSTPDSFVNVTAVGFCVKPANASQYAIVPANGRGIFIQVNGSYIITARDMYMPR 993

Qy 49 RETIEEL-TGTRFDLQQLVNLTLSGHIDEDD-----DEFTLSWS 88
Db 994 AITAGDIVTLTSCQANYVSVNKTIVITTFVDNDDFDNDELSKWN 1038

RESULT 6
US-08-362-240A-15
; Sequence 15, Application US/08362240A
; Patent No. 5965138
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D
; APPLICANT: Junker, David
; APPLICANT: Wild, Martha A
; TITLE OF INVENTION: Recombinant Herpesvirus and Uses Thereof
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White

; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,240A
; FILING DATE: Dec-22-94
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)278-0400
; TELEFAX: (212)391-0526
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1162 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-362-240A-15

Query Match 14.4%; Score 65; DB 2; Length 1162;
Best Local Similarity 26.7%; Pred. No. 6.7;
Matches 28; Conservative 11; Mismatches 46; Indels 20; Gaps 3;

Qy 4 VFIAFQANEDSRPIYDAIVA-----DNPRVVVSPGMVKIDAPDRLTIR 48
Db 934 VFHFSTPDSFVNVTAVGFCVKPANASQYAIVPANGRGIFIQVNGSYIITARDMYMPR 993

Qy 49 RETIEEL-TGTRFDLQQLVNLTLSGHIDEDD-----DEFTLSWS 88
Db 994 AITAGDIVTLTSCQANYVSVNKTIVITTFVDNDDFDNDELSKWN 1038

RESULT 7
PCT-US95-10245-15
; Sequence 15, Application PC/TUS9510245
; GENERAL INFORMATION:
; APPLICANT: SYNTRO CORPORATION
; TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys And Uses Thereof
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10245
; FILING DATE: 09-AUG-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)278-0400
; TELEFAX: (212)391-0526
; TELEX: 422523

INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-10245-15

Query Match 14.4%; Score 65; DB 4; Length 1162;
Best Local Similarity 26.7%; Pred. No. 6.7;
Matches 28; Conservative 11; Mismatches 46; Indels 20; Gaps 3;

QY 4 VFTAFQANEDSRPIVDAIVA-----DNPRVAVVESPGMKIDAPDRLTIR 48
Db 934 VFTHSYTDPDSFVNVTAVGFCVKPANASQYAIVPANGRGIFQVNGSYIITARDMYMR 993
QY 49 RETIEL-TGTRPDLOQLVNLTLTSGHIDEDD-----DEFTLSWS 88
Db 994 AITAGDIVTLTSCQANYSVNKTIVITTFVDNDDDFDNELSKWN 1038

RESULT 8
US-08-851-190-3
Sequence 3, Application US/08851190
Patent No. 6074843

GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Goli, Surya K.
TITLE OF INVENTION: HUMAN TSC-22-LIKE PROTEIN
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,190
FILING DATE: Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0285 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 292 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: TONSNOT01
CLONE: 736663
US-08-851-190-3

Query Match 14.2%; Score 64; DB 3; Length 292;
Best Local Similarity 24.6%; Pred. No. 1.3;
Matches 17; Conservative 13; Mismatches 21; Indels 18; Gaps 2;

QY 13 DSRPIVDAVNPRVAVVESPGMKIDAPDRLTIRRETIEELTGTRFDLOQLVNLTL 72
Db 20 DSRPSPALYFTDASLVHKSP-----DPFGVAAQKFSLAH---SMLAI 61
QY 73 SGHIDEDDD 81
Db 62 SGHLSDSD 70

RESULT 9
5212296-6
Patent No. 5212296
APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH
J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.
TEPPERMAN, JAMES M.
TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
CYTOCHROMES
NUMBER OF SEQUENCES: 19
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/569,781
FILING DATE: 23-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 464,499
FILING DATE: 12-JAN-1990
APPLICATION NUMBER: 405,605
FILING DATE: 11-SEP-1989
SEQ ID NO: 6:
LENGTH: 406
5212296-6

Query Match 13.9%; Score 62.5; DB 5; Length 406;
Best Local Similarity 35.5%; Pred. No. 3.2;
Matches 22; Conservative 9; Mismatches 28; Indels 3; Gaps 3;
QY 19 DAIVADNPR-AVVVESP-GMKIDAPDRLTIRRETIEELTGTRFDLOQLVNLTLSGHI 76
Db 76 DNFATSPRFEAVRESPQAFICGLDPHECTRRMTISEFTVKRIKMRPEVEV-VHGFL 134
QY 77 DE 78
Db 135 DE 136

RESULT 10
US-08-505-218-6
Sequence 6, Application US/08505218
Patent No. 5914447
GENERAL INFORMATION:
APPLICANT: ARAYA, ALEJANDRO
APPLICANT: MOURAS, ARMAND
TITLE OF INVENTION: TRANSGENIC PLANTS INCLUDING A HYBRID
NUCLEIC ACID SEQUENCE, COMPRISING AN UNEDITED
MITOCHONDRIAL GENE FRAGMENT FROM HIGHER PLANTS AND PROCESS
FOR PRODUCING THEM
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: LOWE, PRICE, LEBLANC & BECKER
STREET: 99 Canal Center Plaza, Suite 300
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/505,218
FILING DATE: 03-NOV-1995
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:
NAME: KRAUS, ERIC J
REGISTRATION NUMBER: 36,190
REFERENCE/DOCKET NUMBER: 1169-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 684-1111
TELEFAX: (703) 684-1124
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 322 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-505-218-6

Query Match 13.7%; Score 62; DB 2; Length 322;
Best Local Similarity 32.1%; Pred. No. 2.7;
Matches 27; Conservative 12; Mismatches 25; Indels 20; Gaps 6;
QY 13 DSRPIVDA-----IV--ADNPRAVVVSPGMWKIDA-PDR-----LTIRRETI-----E 53
Db 229 DNRVVVPAKHLRMIVTADVPHSWAVPSSG-VKCDVAPGRSLTFISVOREGVYIGQCS 287
QY 54 ELTGTRFDLQQLQVNLITLSGHID 77
Db 288 EIRGTHAFTPIVVEAVTLKDYAD 311

RESULT 11
US-07-960-389-2
Sequence 2, Application US/07960389
Patent No. 5705611
GENERAL INFORMATION:
APPLICANT: HAYASHIDA, Kasuhiro;
TITLE OF INVENTION: Human GM-CSF Receptor Component
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Schering-Plough Corporation
STREET: 2000 Galloping Hill Road
CITY: Kenilworth
STATE: New Jersey
COUNTRY: USA
ZIP: 07033
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disc
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System Software 7.1
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07960,389
FILING DATE: 07-JAN-1993
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 554,745
FILING DATE: 18-JUL-1990
APPLICATION NUMBER: PCT/US 91/04846
FILING DATE: 16-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Blasdale, John H. C.
REGISTRATION NUMBER: 31,895
REFERENCE/DOCKET NUMBER: DX0143Q
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 298-2902
TELEFAX: (908) 298-5388
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 897 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:

OTHER INFORMATION: Human GM-CSF receptor; Signal Sequence: -17 to -1; Tran
US-07-960-389-2
Query Match 13.7%; Score 62; DB 1; Length 897;
Best Local Similarity 24.6%; Pred. No. 11;
Matches 28; Conservative 9; Mismatches 31; Indels 46; Gaps 5;
QY 3 NVFTAFQANEDSRPIVDAIVADN-----PRAVVVSPGMWKIDA-----PDR 44
Db 58 NVTILRRVNEDLLFVSCDLSDDMPWSACPHRCVPRCVVTVDYFSFQDR 117
QY 45 LTIRRETIETGTGRFDLQQLQVNLITLSGH-----IDEDDDFTLSWS 88
Db 118 -----PLGTRL-----TVILTQHVQPPPRDLQISTDDQDFILTWS 153

RESULT 12
US-08-906-769-89
Sequence 89, Application US/08906769
Patent No. 6077687
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/906,769
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 89:
SEQUENCE CHARACTERISTICS:
LENGTH: 256 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-906-769-89

Query Match 13.5%; Score 61; DB 3; Length 256;
Best Local Similarity 22.9%; Pred. No. 2.6;
Matches 16; Conservative 16; Mismatches 26; Indels 12; Gaps 1;
QY 25 NPRAVVVESP-----GMVKIDAPDRLTIRRETIETGTGRFDLQQLQVNLITL 72

```
Db 90 HPXAQIIKHPAYGNVTDIMEXALIKVRRPRLNRTVTKLTDVGKMPSGELATVTG 149
QY 73 SGHIEDDEDE 82
Db 150 WGNLGEDEDD 159

RESULT 13
PCT-US95-14442A-89
; Sequence 89, Application PC/TUS9514442A
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA
; TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/14442A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gary J. Connell
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 256 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-14442A-89

Query Match 13.5%; Score 61; DB 4; Length 256;
Best Local Similarity 22.9%; Pred. No. 2.6;
Matches 16; Conservative 16; Mismatches 26; Indels 12; Gaps 1;

QY 25 NPRAVVVEP-----GMVKIDAPDLTIRRETIETELGTFRDLQQLQVNLITL 72
Db 90 HPXAQIIKHPAYGNVTDIMEXALIKVRRPRLNRTVTKLTDVGKMPSGELATVTG 149
QY 73 SGHIEDDEDE 82
Db 150 WGNLGEDEDD 159

RESULT 14
US-08-266-451B-22
; Sequence 22, Application US/08266451B
```

```
; Patent No. 5623054
; GENERAL INFORMATION:
; APPLICANT: Zhang et al.
; TITLE OF INVENTION: CRUCIFER APT PROTEINS AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/266,451B
; FILING DATE: 23-June-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Lech, Karen F.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/219001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 339
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-266-451B-22

Query Match 13.2%; Score 59.5; DB 1; Length 339;
Best Local Similarity 25.6%; Pred. No. 6.2;
Matches 23; Conservative 14; Mismatches 30; Indels 23; Gaps 3;

QY 15 RPVIDAIVADNPRAVV-----ESPGMVKIDAPDLRTI-----RRETI 52
Db 159 KPILEDIDAGGSPAMKYWNPDEVLKKGLEANGPVACLPDQTVSAEPEVAEEGEEESI 218
QY 53 BELGTFRDLQQLQVNLITLSGHIEDDEDE 82
Db 219 VHQATSLGDVGLKAAALAS-GGNKDEEDSE 247

RESULT 15
US-08-748-725-22
; Sequence 22, Application US/08748725
; Patent No. 5859346
; GENERAL INFORMATION:
; APPLICANT: Zhang et al.
; TITLE OF INVENTION: CRUCIFER APT PROTEINS AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/748,725
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/266,451
; FILING DATE: 23-June-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lech, Karen F.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/219002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 339
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-748-725-22

Query Match 13.2%; Score 59.5; DB 2; Length 339;
Best Local Similarity 25.6%; Pred. No. 6.2;
Matches 23; Conservative 14; Mismatches 30; Indels 23; Gaps 3;
QY 15 RPIVDAIVADNPRAVV-----ESPGMWKIDAPDRLTI-----RRETI 52
Db 159 KPILDEIDAGGSAAMKYWNPVVKLKGAMGMPVAGLPDQTVSAEPEVAEGEEREESI 218
QY 53 EELGTREDFLQQLQVNLITLSGHIDEDDE 82
Db 219 VHQFASLGDEGLKALAS-GGNKDEDESE 247

Search completed: September 26, 2000, 20:22:40
Job time: 3814 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:24:12 ; Search time 82.1 Seconds
(without alignments)
67.086 Million cell updates/sec

Title: US-09-430-029-4

Perfect score: 451

Sequence: 1 MSNVIFAQANEDSRPIVDA.....ITLSGHIDDDDEFTLSWSH 89

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_64:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	247	54.8	90	1 S44305	phenol 2-monooxyge
2	247	54.8	90	1 C37831	phenol 2-monooxyge
3	232	51.4	89	1 S47289	phenol 2-monooxyge
4	83.5	18.5	862	2 T36380	probable large ATP
5	77	17.1	1684	2 T02632	hypothetical prote
6	70.5	15.6	174	2 G69411	small heat shock p
7	69	15.3	822	2 S77112	DNA mismatch repal
8	68.5	15.2	143	2 D75617	response regulator
9	67.5	15.0	410	2 J42887	cytochrome P450 sc
10	66.5	14.7	2277	2 B47648	perlecan homolog u
11	66.5	14.7	2295	2 T19820	hypothetical prote
12	66.5	14.7	2482	2 T19819	hypothetical prote
13	66.5	14.7	2482	2 A47648	perlecan homolog u
14	66.5	14.7	3375	2 T19821	hypothetical prote
15	66	14.6	138	2 D39049	methane monooxygen
16	66	14.6	181	2 A40607	monofunctional cho
17	66	14.6	294	2 D81430	UDP-3-O-[3-hydroxy
18	66	14.6	390	2 C69276	coenzyme F420--qui
19	66	14.6	619	2 A43361	Ets-related transc
20	66	14.6	649	2 S74823	N-acetylmutamoyl-L
21	65	14.4	1162	2 S07421	E2 glycoprotein pr
22	65	14.4	1232	2 S40766	hypothetical prote
23	64.5	14.3	316	2 B47157	magnesium transpor
24	64.5	14.3	452	2 I49595	cytokeratin 15 - m
25	64	14.2	347	2 A75537	hypothetical prote
26	64	14.2	487	1 BWSOGM	gtfA protein - Str
27	64	14.2	1162	2 S14939	E2 glycoprotein pr
28	63.5	14.1	224	2 G69783	transcription regu
29	63.5	14.1	394	2 T06180	methionine adenosy

30	63.5	14.1	420	1 DCECD	diaminopimelate de
31	63	14.0	366	2 S66351	methionine adenosy
32	63	14.0	557	2 S21596	extracellular prot
33	63	14.0	828	2 S56250	probable membrane
34	62.5	13.9	139	2 A69052	riboflavin synthas
35	62.5	13.9	205	2 S39863	late competence op
36	62.5	13.9	360	2 T06592	methionine adenosy
37	62.5	13.9	406	2 A35401	cytochrome p450 10
38	62.5	13.9	964	2 T21865	hypothetical prote
39	62.5	13.9	1541	2 T30227	pipecolate-incorpo
40	62.5	13.9	4077	3 T17484	hypothetical prote
41	62	13.7	80	2 D32338	dexA.1 protein - p
42	62	13.7	166	2 I69005	histocompatibility
43	62	13.7	260	1 OBWT2	cytochrome-c oxida
44	62	13.7	260	2 S54306	cytochrome-c oxida
45	62	13.7	260	2 T01704	cytochrome-c oxida

ALIGNMENTS

RESULT 1

S44305

phenol 2-monooxygenase (EC 1.14.13.7) component M - Pseudomonas putida

N;Alternate names: phenolhydroxylase chain C

C;Species: Pseudomonas putida

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C;Accession: D58972; S54766; S44305; S47416

R;Ng, L.C.; Shingler, V.; Size, C.C.; Poh, C.L.

Gene 151, 29-36, 1994

A;Title: Cloning and sequences of the first eight genes of the chromosomally encoded

A;Reference number: A58972; MUID:95129877

A;Accession: D58972

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-90 <NGL>

A;Cross-references: EMBL:X79063; NID:q483477; PIDN:CAA55662.1; PID:q483480

A;Experimental source: strain P35X (NCBI 9869)

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1994

R;Herrmann, H.; Mueller, C.; Schmidt, I.; Mahnke, J.; Petruschka, L.; Hahnke, K.

Mol. Gen. Genet. 247, 240-246, 1995

A;Title: Localization and organization of phenol degradation genes of Pseudomonas put

A;Reference number: S54761; MUID:95272534

A;Accession: S54766

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-41,'V',43-55,'K',57-76,'V',78-90,'N' <HE2>

A;Cross-references: EMBL:X80765; NID:g527546; PIDN:CAA56742.1; PID:g527549

A;Experimental source: strain H

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994

C;Genetics:

A;Gene: phlC; phhM

C;Superfamily: phenol 2-monooxygenase component M

C;Keywords: oxidoreductase

Query Match

Best Local Similarity 54.8%; Score 247; DB 1; Length 90;

Matches 46; Conservative 20; Mismatches 22; Indels 0; Gaps 0;

Qy 2 SNVFIAQANEDSRPIVDAIVADNPRAVVVSPGMVKIDAPDLRTIRRETEELTGTRFD 61

Db 3 SLVVFIAQDNDAVYLAIQDNPFAVQVHHFAMIRAEKRLRETRVEENLGRWD 62

Qy 62 LQLQVNLITLSGHIDDDDEFTLSWSH 89

Db 63 VQEMLVDTIIGGNIDDDDFVLEWN 90

RESULT 2

C37831

phenol 2-monooxygenase (EC 1.14.13.7) chain P2 - Pseudomonas sp. (strain CF600)

C;Species: Pseudomonas sp.

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

A:Accession: C37831
R:Nordlund, I.; Powlowski, J.; Shingler, V.
J. Bacteriol. 172, 6826-6833, 1990
A:Title: Complete nucleotide sequence and polypeptide analysis of multicomponent phenol
A:Reference number: A37831; MUID:91072230
A:Accession: C37831
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-90 <NOR>
A:Cross-references: GB:M60276; GB:M37764; NID:g151449; PIDN:AAA25941.1; PID:g151452
C:Superfamily: phenol 2-monooxygenase component M
C:Keywords: oxidoreductase

Query Match 54.8%; Score 247; DB 1; Length 90;
Best Local Similarity 51.1%; Pred. No. 8.2e-19;
Matches 45; Conservative 21; Mismatches 22; Indels 0; Gaps 0;

QY 2 SNVTFAGNEDSRPDAIVADNPRAVVGSPGMVKIDAPDLRTIRRETEELTGTRFD 61

DB 3 SLVYAFQDNQARVAVVETAIQDNPHAVVQHHPAMIRAEKREIRRETEELTGTRFD 62

QY 62 LQQLQVNLITLSGHIDDDDEFTLSWSH 89

DB 63 VQELVDVITIGGNVDEDDREVLKMN 90

RESULT 3
S47289
phenol 2-monooxygenase (EC 1.14.13.17) chain mopM - Acinetobacter calcoaceticus
N:Alternate names: phenol hydroxylase
C:Species: Acinetobacter calcoaceticus
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S70082; S47289
R:Ehrt, S.; Schirmer, F.; Hillen, W.
Mol. Microbiol. 18, 13-20, 1995
A:Title: Genetic organization, nucleotide sequence and regulation of expression of genes
A:Reference number: S70080; MUID:96154937
A:Accession: S70082
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-89 <EH2>
A:Cross-references: EMBL:Z36909; NID:g535279; PIDN:CAA85382.1; PID:g535282
A:Experimental source: strain NCIB8250
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1994
C:Genetics:
A:Gene: mopM
C:Superfamily: phenol 2-monooxygenase component M
C:Keywords: aromatic hydrocarbon catabolism; oxidoreductase

Query Match 51.4%; Score 232; DB 1; Length 89;
Best Local Similarity 46.0%; Pred. No. 2.9e-17;
Matches 40; Conservative 21; Mismatches 26; Indels 0; Gaps 0;

QY 2 SNVTFAGNEDSRPDAIVADNPRAVVGSPGMVKIDAPDLRTIRRETEELTGTRFD 61

DB 3 SKVYALQNDTSRVIIIEIQDNPEATIQYLPAMIRVETGELVVRVETVSEKLQGNWD 62

QY 62 LQQLQVNLITLSGHIDDDDEFTLSWS 88

DB 63 IQELQNLITLGGNVDEDDSFLLKWN 89

RESULT 4
T36380
probable large ATP-binding protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36380
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, April 1999

A:Reference number: Z21573

A:Accession: T36380
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-862 <OL1>
A:Cross-references: EMBL:AL049628; PIDN:CAB40869.1; GSPDB:GN00070; SCOEDB:SCE94.20
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCE94.20

Query Match 18.5%; Score 83.5; DB 2; Length 862;
Best Local Similarity 28.6%; Pred. No. 0.85;
Matches 18; Conservative 17; Mismatches 27; Indels 1; Gaps 1;

QY 27 RAVVSPGMVKIDAPDLRTIRRETEELTGTRFDLQQLQVNLITLSGHIDDDDEFTLS 86

DB 556 RALLRS-GLLRPAEGRVDFLHRTFDYLGARLAVQEMDFLLVNHHLDEWDDVILLA 614

QY 87 WSH 89

DB 615 LAH 617

RESULT 5
T02632
hypothetical protein DJ1186C01.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
C:Accession: T02632
R:Bemis, G.; Rohlfing, T.; Morris, M.
submitted to the EMBL Data Library, July 1998
A:Description: The sequence of Homo sapiens PAC clone DJ1186C01.
A:Reference number: Z14682
A:Accession: T02632
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1684 <BEM>
A:Cross-references: EMBL:AC004991; NID:g3342751; PIDN:AC27675.1; PID:g3342752
C:Genetics:
A:Map position: 7
A:Note: intron positions not resolved (incomplete sequence)
A:Note: WUGSC:H_DJ1186C01.1

Query Match 17.18%; Score 77; DB 2; Length 1684;
Best Local Similarity 26.7%; Pred. No. 8.6;
Matches 24; Conservative 20; Mismatches 24; Indels 22; Gaps 4;

QY 2 SNVTFAGNEDSRPDAIVADNPRAVVGSPG-----MVKIDAPDLRTIRRETEELTG 57

DB 1250 SRPMSFLKHKDRPLRDVLAANPNRFTLLPFGAQTAVRPGSPSTSTM----- 1299

QY 58 TRFDLQQLQVNLITLSGHIDDDDEFTLSW 87

DB 1300 -RLDLQQAIIISI---IVKNDQ----SW 1321

RESULT 6
G69411
small heat shock protein (hsp20-1) homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998
C:Accession: G69411
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiaich, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A:Reference number: A69250; MUID:98049343
A:Accession: G69411

C;Accession: T19819; T27488
R;Baynes, C.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z19182
A;Accession: T19819
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2482 <W1>
A;Cross-references: EMBL:Z93375; PIDN:CAB07567.1; GSPDB:GN00020; CESP:ZC101.2a
A;Experimental source: clone C38C6
R;Percy, C.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z20375
A;Accession: T27488
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2482 <W12>
A;Cross-references: EMBL:Z93395; PIDN:CAB07706.1; GSPDB:GN00020; CESP:ZC101.2a
A;Experimental source: clone ZC101
C;Genetics:
A;Gene: CESP:ZC101.2a
A;Map position: 2
A;Introns: 32/1; 134/1; 225/1; 335/2; 450/3; 739/3; 830/3; 860/2; 1064/2; 1129/1; 115

```

Query Match          14.7%; Score 66.5; DB 2; Length 2482;
Best Local Similarity 28.4%; Pred. No. 1.16e+02;
Matches 21; Conservative 14; Mismatches 24; Indels 15; Gaps 3;

QY 16 PIVDAIVADNPRAVVVSPCMWIDAPDLRTIRRETEELTGTFRDLQ-----QLQVN 68
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1578 PTTNPVSPNPARVIVKSPIRPIIDFAE-----QTVP-E-GSPFKIRCYVPGHPVSQLT 1629
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 69 LITLSGHIDEEDDE 82
      : | | : | | | | |
Db 1630 FRVVSQOLNEDAE 1643
      : | | : | | | | |

RESULT 13
A47648
perlecan homolog unc-52, form 1 precursor - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 16-Jul-1999
C:Accession: A47648
R:Rogalski, T.M.; Williams, B.D.; Mullen, G.P.; Moerman, D.G.
Genes Dev. 7, 1471-1484, 1993
A:Title: Products of the unc-52 gene in Caenorhabditis elegans are homologous to the
A:Reference number: A47648; MUID:93339574
A:Accession: A47648
A>Status: preliminary; nucleic acid sequence not shown

```

```
C:keywords: extracellular matrix
F:149-183/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:190-224/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:233-268/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F:955-1002/Domain: laminin-type EGF-like homology <LEG1>
F:1011-1058/Domain: laminin-type EGF-like homology <LEG2>

Query Match          14.7%;   Score 66.5;   DB 2;   Length 2482;
Best Local Similarity 28.4%;   Pred. No. 1.6e+02;
Matches 21; Conservative 14; Mismatches 24; Indels 15; Gaps 3;

QY      16 PIVDAIVADNPRAVVVYESGMWKIDAPDRLTIIRREIIELTGRFDLQ-----QLQVN 68
       | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
Db     1578 PTTNREVSNPARVIKVSPIRIPIDAE-----QTVPV--GSPFKIRCVPGHPSVQLT 1629
```

```

Db      1578  PTTNRPVESNPARVIKSPIRPIIDPAE-----QTUPE--GSPFKIRCVPVGHPSVOLT 1629
QY      69   LITLSGHIDEDDE 82
Db      1630  FRRVSGQLNEADE 1643

```


RESULT 14

T19821

hypothetical protein ZC101.2e - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999

C:Accession: T19821; T27490

R:Baynes, C.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z19182

A:Accession: T19821

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-3375 <WIL>

A:Cross-references: EMBL:Z93375; PIDN:CAB07569.1; GSPDB:GN00020; CESP:ZC101.2e

A:Experimental source: clone C38C6

R:Percy, C.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z20375

A:Accession: T27490

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-3375 <WIL>

A:Cross-references: EMBL:Z93395; PIDN:CAB07708.1; GSPDB:GN00020; CESP:ZC101.2e

A:Experimental source: clone ZC101

C:Genetics:

A:Gene: CESP:ZC101.2e

A:Map position: 2

A:Introns: 32/1; 134/1; 225/1; 335/2; 450/3; 739/3; 830/3; 860/2; 1064/2; 1129/1; 1158/3; 2613/1; 2684/1; 2757/1; 2813/3; 2863/1; 2900/3; 3084/1; 3176/1; 3250/2

Query Match

Best Local Similarity 14.7%; Score 66.5; DB 2; Length 3375;

Matches 21; Conservative 14; Mismatches 24; Indels 15; Gaps 3;

QY 16

PVTDAIVADNPRAVVGSPGMVKIDAPDLRTIRRETTELTGTRFDLQ-----QLQVN 68

DB 1578 PTTNRVPESNPARIYKSPIRPIIDPAE-----QIVPE--GSPFKIRCIVPGHPVSOLT 1629

QY 69

LITLSGHIDEDEDD 82

DB 1630 FRRVSGQLNEADE 1643

RESULT 15

D39049

methane monooxygenase (EC 1.14.13.25) component B - Methylosinus trichosporium (strain C

C:Species: Methylosinus trichosporium

C>Date: 23-Aug-1991 #sequence_revision 26-Apr-1996 #text_change 08-Oct-1999

C:Accession: S15209; D39049

R:Cardy, D.L.N.; Laidler, V.; Salmond, G.P.C.; Murrell, J.C.

Mol. Microbiol. 5, 335-342, 1991

A:Title: Molecular analysis of the methane monooxygenase (MMO) gene cluster of Methylosin

A:Reference number: S15207; MUID:91251762

A:Accession: S15209

A:Molecule type: DNA

A:Residues: 1-138 <CAR>

A:Cross-references: EMBL:X55394; NID:g44613; PIDN:CAA39070.1; PID:g44616

R:Fox, B.G.; Liu, Y.; Dege, J.E.; Lipscomb, J.D.

J. Biol. Chem. 266, 540-550, 1991

A:Title: Complex formation between the protein components of methane monooxygenase from

A:Reference number: A39049; MUID:91093180

A:Accession: D39049

A>Status: preliminary

A:Molecule type: protein

A:Residues: 2-32 <FOX>

C:Genetics:

A:Gene: mmob

C:Keywords: oxidoreductase

Query Match 14.6%; Score 66; DB 2; Length 138;

Best Local Similarity 25.5%; Pred. No. 6.6;

Matches 24; Conservative 12; Mismatches 46; Indels 12; Gaps 2;

QY 2

SNVFIAFOANEDSRPIVDAIVAD-----NPRVVVSPGMVKIDAPDLRTIRRETI 52

DB 35 SNAVVLVLMKSDE---IDAIIEDIVLKGKAKNPSIVVEDKAGFWIKADGATEIDAAEA 91

QY 53

EELTGTRFDLQQLQOVNLTLSGHIDEDEDDFTLS 86

DB 92 GELLGKPFVYDLLINVSSTVGRAYTLGTRKETIT 125

Search completed: September 26, 2000, 20:24:16

Job time: 3640 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein ~ protein search, using sw model

Run on: September 26, 2000, 20:22:44 ; Search time 60.65 Seconds
(without alignments)
29.824 Million cell updates/sec

Title: US-09-430-029-6

Perfect score: 666

Sequence: 1 MAVIALKPYDFPVKDAVEK.....HKDLISFRTPGLDGLGASG 118

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues

Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/PCUS_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	370	55.6	93	1	US-08-319-387-5
2	78.5	11.8	495	1	US-07-841-997A-2
3	78.5	11.8	495	1	US-08-290-301-2
4	78.5	11.8	495	2	US-08-588-983-2
5	78.5	11.8	495	2	US-08-588-976-2
6	71.5	10.7	273	1	US-08-221-750A-9
7	62	9.3	655	2	US-08-272-255-18
8	62	9.3	655	4	PCT-US95-08565-18
9	61	9.2	336	2	US-08-987-080-156
10	61	9.2	336	2	US-08-997-362-156
11	61	9.2	495	2	US-09-132-619-2
12	61	9.2	495	3	US-09-282-803B-2
13	61	9.2	541	2	US-09-132-619-6
14	61	9.2	541	3	US-09-282-803B-6
15	59.5	8.9	495	2	US-08-749-902-3
16	59.5	8.9	500	2	US-09-132-619-8
17	59.5	8.9	500	3	US-09-282-803B-8
18	59.5	8.9	681	2	US-08-272-255-6
19	59.5	8.9	681	2	US-08-272-255-9
20	59.5	8.9	681	4	PCT-US95-08565-6
21	59.5	8.9	681	4	PCT-US95-08565-9
22	59.5	8.9	726	2	US-08-313-185-49
23	59.5	8.9	726	2	US-08-459-499-13
24	59	8.9	267	1	US-07-928-611-13
25	59	8.9	267	2	US-08-487-811A-13
26	59	8.9	267	4	PCT-US93-07370-13
27	59	8.9	467	1	US-08-056-051-6
28	59	8.9	467	1	US-07-928-611-22

29 59 8.9 467 2 US-08-487-811A-22 Sequence 22, Appl
30 59 8.9 467 4 PCT-US93-07370-22 Sequence 22, Appl
31 59 8.9 542 1 US-07-814-964-13 Sequence 13, Appl
32 59 8.9 542 1 US-08-258-442-13 Sequence 13, Appl
33 59 8.9 542 1 US-08-328-809-8 Sequence 8, Appl
34 59 8.9 542 4 PCT-US92-11107-13 Sequence 13, Appl
35 59 8.9 558 4 PCT-US93-03027-2 Sequence 2, Appl
36 58.5 8.8 360 1 US-08-205-506A-2 Sequence 2, Appl
37 58.5 8.8 360 4 PCT-US94-02389-2 Sequence 2, Appl
38 58 8.7 482 1 US-07-792-885A-1 Sequence 1, Appl
39 58 8.7 482 1 US-08-142-439A-7 Sequence 7, Appl
40 58 8.7 482 2 US-08-869-477-7 Sequence 7, Appl
41 58 8.7 549 1 US-08-494-168-8 Sequence 8, Appl
42 58 8.7 1298 2 US-08-690-473-2 Sequence 2, Appl
43 58 8.7 1319 2 US-08-290-731C-2 Sequence 2, Appl
44 58 8.7 1336 2 US-08-290-731C-6 Sequence 6, Appl
45 57.5 8.6 902 1 US-08-396-479B-6 Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-319-387-5
; Sequence 5, Application US/08319387
; Patent No. 5543317
; GENERAL INFORMATION:
; APPLICANT: Shields, Malcolm S.
; APPLICANT: Francesconi, Stephen C.
; TITLE OF INVENTION: Microbial Degradation of Trichloroethylene,
; TITLE OF INVENTION: Dichloroethylenes and Aromatic Pollutants
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/319,387
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/167,457
; FILING DATE: 15-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/694,718
; FILING DATE: 02-MAY-1991
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UWF-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-319-387-5

QY	14	KUAVENFAPLILFVWENHLMFAPECLFPLPDMF	GALAGVOLEPVGTHFQFARLWD	73
Db	80	KOLLEKFNVPJFLSDE	-----IADLHYNGFNSILMPLFHYHB	-GEINPD 124
QY	74	RVWEF----	RSGEWPAPDPKSLAGNGLGHKDLI	103
Db	125	ENAWFGYNEANOTFTNEIAKT	-----MHNHNDLI	152


```

Query Match          9.38; Score 62; DB 4; Length 655;
Best Local Similarity 39.6%; Pred. No. 39;
Matches 21; Conservative 4; Mismatches 12; Indels 16;

      QY 17 VEXFAPLLYVCWB-----NHLMEFAPF---CLPLPDPMPFGALAGDVLPP 59
          | : | | | | | | : | | | | | | | | | | | | | | |
      IdB 20 VQSFNADLLYEPVEVDGQSFTEMF-APFWNRCLSLMPYDP-----AAPLLPP 66

RESULT          9
US-08-997-080-156
; Sequence 156, Application US/08997080
; Patent No. 5968524
; GENERAL INFORMATION:
; APPLICANT: WATSON, JAMES D.
; APPLICANT: TAN, PAUL L.J.
; TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman

```

STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
APPLICATION DATA:
APPLICATION NUMBER: US/08/997,080
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1007

TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:

INFORMATION FOR SEQ ID NO: 156:
SEQUENCE CHARACTERISTICS:
LENGTH: 336 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-997-080-156

Query Match 9.2%; Score 61; DB 2; Length 336;
Best Local Similarity 23.0%; Pred. No. 22;
Matches 28; Conservative 10; Mismatches 28; Indels 56; Gaps 8;

QY 36 PAPCLP-LPPDMPFAGALAGDVLPPVGYHDPFAKI-DWDRV-----EWFR 79
| | | | | : : : : : | | | | |
Db 19 PSLCSPALAPS--HAGLDNELSLGVHGGQPEHLTIQWDTFLNGVFPFLDRNRLTREW 76
| | | | | : : : : : | | | | |
QY 80 SGE-----PWAPDPAKSLAGNGLGHKDLISRTPTGL--DG 112
| | | | | : : : : : | | | | |
Db 77 SGKATYVVAGEGADEFECTLELGYHVGFPWS-----LGVGIN----FSYTTNPITVDG 125
| | | | | : : : : : | | | | |
QY 113 LG 114
| | | | | : : : : : | | | | |
Db 126 YG 127

RESULT 10
US-08-997-362-156
Sequence 156, Application US/08997362
Patent No. 5985287
GENERAL INFORMATION:
APPLICANT: Tan, Paul
APPLICANT: Hiyama, Jun
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Scott, Linda
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
NUMBER OF SEQUENCES: 194
CORRESPONDENCE ADDRES:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/997,362
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
FILING DATE: June 12, 1997
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
FILING DATE: August 29, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002c2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 156:
SEQUENCE CHARACTERISTICS:
LENGTH: 336 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-997-362-156

Query Match 9.2%; Score 61; DB 2; Length 336;
Best Local Similarity 23.0%; Pred. No. 22;
Matches 28; Conservative 10; Mismatches 28; Indels 56; Gaps 8;

QY 36 PAPCLP-LPPDMPFAGALAGDVLPPVGYHDPFAKI-DWDRV-----EWFR 79
| | | | | : : : : : | | | | |
Db 19 PSLCSPALAPS--HAGLDNELSLGVHGGQPEHLTIQWDTFLNGVFPFLDRNRLTREW 76
| | | | | : : : : : | | | | |
QY 80 SGE-----PWAPDPAKSLAGNGLGHKDLISRTPTGL--DG 112
| | | | | : : : : : | | | | |
Db 77 SGKATYVVAGEGADEFECTLELGYHVGFPWS-----LGVGIN----FSYTTNPITVDG 125
| | | | | : : : : : | | | | |
QY 113 LG 114
| | | | | : : : : : | | | | |
Db 126 YG 127

RESULT 11
US-09-132-619-2
Sequence 2, Application US/09132619B
Patent No. 5958697
GENERAL INFORMATION:
APPLICANT: Shan, Bei
APPLICANT: Nitta, Masahiro
TITLE OF INVENTION: CYP7 Promoter-Binding Factors
FILE REFERENCE: T97-013
CURRENT APPLICATION NUMBER: US/09/132,619B
CURRENT FILING DATE: 1998-08-11
EARLIER APPLICATION NUMBER: 60/067,708
EARLIER FILING DATE: 1997-12-08
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn ver. 2.0
SEQ ID NO 2
LENGTH: 495
TYPE: PRT
ORGANISM: human
US-09-132-619-2

Query Match 9.2%; Score 61; DB 2; Length 495;
Best Local Similarity 22.3%; Pred. No. 36;
Matches 27; Conservative 17; Mismatches 37; Indels 40; Gaps 5;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:24:21 ; Search time 82.1 seconds
(without alignments)
88.945 Million cell updates/sec

Title: US-09-430-029-6
Perfect score: 666
Sequence: 1 MAVIALKPYDFPVKDAVEKF.....HKDLISFRTPLDGLGGASF 118

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_64: *
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	215.5	32.4	119	1 E37831	phenol 2-monooxyge
2	212.5	31.9	119	1 S44307	phenol 2-monooxyge
3	143	21.5	120	1 S47291	phenol 2-monooxyge
4	127	19.1	671	2 A45730	phenol 2-monooxyge
5	84.5	12.7	271	2 S46295	chlorophyll a/b-bi
6	78.5	11.8	487	1 S52261	NADH dehydrogenase
7	72.5	10.9	495	2 S34979	alpha, alpha-trehal
8	71.5	10.7	273	2 D47301	VirB9 homolog - Bo
9	71.5	10.7	640	2 S69546	phosphoenolpyruvat
10	71	10.7	522	2 S52216	viral proteinase -
11	68	10.2	201	2 S26045	sex-determining pr
12	68	10.2	482	2 T40809	conserved hypotet
13	67.5	10.1	372	2 S60207	fomP protein - Str
14	66.5	10.0	94	2 S61459	heat shock transcr
15	66.5	10.0	102	2 G25035	hypothetical prote
16	66.5	10.0	345	2 S12788	transcription fact
17	66.5	10.0	476	2 T43464	hypothetical prote
18	66.5	10.0	691	2 T46476	hypothetical prote
19	66	9.9	405	2 T16836	proline-rich prote
20	66	9.9	448	2 T06076	probable cell divi
21	66	9.9	980	2 G75523	diaphanous-related
22	66	9.9	1171	2 T17454	glycerolaldehyde-3-p
23	65.5	9.8	501	2 T44939	beta-fructofuranos
24	65.5	9.8	562	2 S57951	ropy-2 protein [im
25	65.5	9.8	710	2 T46589	interleukin-11 rec
26	65	9.8	422	2 I37891	paired box transcr
27	65	9.8	442	2 A57374	NADH dehydrogenase
28	65	9.8	464	1 JE0092	isobutyryl-CoA mut
29	65	9.8	566	1 T43706	

30 64.5 9.7 131 2 B75433
31 64.5 9.7 339 2 S39979
32 64.5 9.7 340 2 S40414
33 64.5 9.7 396 2 S68772
34 64.5 9.7 396 2 S68773
35 64.5 9.7 437 2 S15704
36 64 9.6 169 2 F70799
37 64 9.6 342 2 S57612
38 64 9.6 464 1 A39362
39 64 9.6 472 2 T33800
40 64 9.6 566 1 T46549
41 64 9.6 622 2 H64447
42 64 9.6 690 2 T31487
43 64 9.6 705 1 C1HURB
44 64 9.6 990 2 T14756
45 64 9.6 1182 2 I48378

ALIGNMENTS

RESULT 1

E37831
phenol 2-monooxygenase (EC 1.14.13.7) chain P4 - Pseudomonas sp. (strain CF600)
C:Species: Pseudomonas sp.
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: E37831
R:Nordlund, I.; Powlowski, J.; Shingler, V.
J. Bacteriol. 172, 6826-6833, 1990
A:Title: Complete nucleotide sequence and polypeptide analysis of multicomponent phenol 2-monooxygenase component O
A:Reference number: A37831; MUID: 91072230
A:Accession: E37831
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-119 <NOR>
A:Cross-references: GB:M60276; GB:M37764; NID:g151449; PIDN:AAA25943.1; PID:g151454
C:Superfamily: phenol 2-monooxygenase component O
C:Keywords: Oxidoreductase

Query Match 32.4%; Score 215.5; DB 1; Length 119;
Best Local Similarity 37.6%; Pred. No. 1.1e-15;
Matches 44; Conservative 18; Mismatches 54; Indels 1; Gaps 1;

QY 1 MAVIALKPYDFPVKDAVEKPPA-PLLYVCVENHLMFPAPCLPLPDMPFGALAGDVLPP 59
Db 1 MTVNSIGETATPRDVOQANFGMQLLYLWEEHLMYCSALAFVAPGMPFAEFLEQLKRP 60
QY 60 VGYGHPDFAKIDWRVWFVRSRSGEPWAPDPKASLAGNGLGHKDLISFRTPLDGLGGA 116
Db 61 ATHAHSDSAKIDFSQALWQLNDQPTFPDYAASLEANGIDHKSMLRLNTPGLNGIQGS 117

RESULT 2

S44307
phenol 2-monooxygenase (EC 1.14.13.7) component O - Pseudomonas putida
N:Alternate names: phenolhydroxylase chain E
C:Species: Pseudomonas putida
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: F58972; S54764; S44307; S47418
R:Ng, L.C.; Shingler, V.; Size, C.C.; Poh, C.L.
Gene 151, 29-36, 1994
A:Title: Cloning and sequences of the first eight genes of the chromosomally encoded
A:Reference number: A58972; MUID: 95129877
A:Accession: F58972
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-119 <NGL>
A:Cross-references: EMBL:X79063; NID:g483477; PIDN:CAA55664.1; PID:g483482
A:Experimental source: strain P35X (NCBI 9869)
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1994
R:Herrmann, H.; Mueller, C.; Schmidt, I.; Mahnke, J.; Petruschka, L.; Hahnke, K.
Mol. Gen. Genet. 247, 240-246, 1995

Query Match 32.4%; Score 215.5; DB 1; Length 119;
Best Local Similarity 37.6%; Pred. No. 1.1e-15;
Matches 44; Conservative 18; Mismatches 54; Indels 1; Gaps 1;

QY 1 MAVIALKPYDFPVKDAVEKPPA-PLLYVCVENHLMFPAPCLPLPDMPFGALAGDVLPP 59
Db 1 MTVNSIGETATPRDVOQANFGMQLLYLWEEHLMYCSALAFVAPGMPFAEFLEQLKRP 60
QY 60 VGYGHPDFAKIDWRVWFVRSRSGEPWAPDPKASLAGNGLGHKDLISFRTPLDGLGGA 116
Db 61 ATHAHSDSAKIDFSQALWQLNDQPTFPDYAASLEANGIDHKSMLRLNTPGLNGIQGS 117

A:Title: Complete nucleotide sequence of tbuD, the gene encoding phenol/cresol hydroxylase
A:Reference number: A45730; MUID:93015703
A:Accession: A45730
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-671 <KUK>
A:Cross-references: GB:M98806; NID:g151573; PIDN:AAA25992.1; PTD:g151574
C:Genetics:
A:Gene: tbuD
C:Keywords: flavoprotein; oxidoreductase

Query Match	19.1%	Score 127;	DB 2;	Length 671;
Best Local Similarity	27.7%	Pred. No. 1.5e-05;		
Matches	41;	Conservative 13;	Mismatches 28;	Indels

Qy	1	MAVIA-LK-PYDFPVKDAVEKFPAPL-----	LYVCWEN
		: :	: :
Db	202	MSVTNLKAAAPYKFLMKVDENFPGLSTSGGKTTISAD	SAIDAALHAWD
		: :	: :
Qy	38	PFCLPLPDMFPGALAGDVLPPVGYVHPDFAK----	IDMDRVSEFSGSEFPAPD
			: :
Db	257	-----DLGA-----	AWHLDEASCLRAVDNNAQWFKSGQPWTFD
			: :
Qy	92	-----LAGNGLGHKDLISFRTPLDGLG	114
			: :
Db	297	EGRFVFLAGDA-----	RHRHPLLTGIG 317
			: :

RESULT 5
S46295 chlorophyll a/b-binding protein type II - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 01-Feb-1995 #sequence_revision 19-Jan-1996 #text_change 26-Aug-1996
C:Accession: S46295
R:Zhang, H.; Wang, J.; Goodman, H.M.
Plant Mol. Biol. 25, 551-557, 1994
A:Title: Differential expression in Arabidopsis of Lhca2, a PSI cab gene.
A:Reference number: S46295; MUID:94325479
A:Accession: S46295
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-271 <ZHA>
A:Cross-references: EMBL:U03395; NID:g541564; PIDN:AAA57542.1; PID:g541564
C:Genetics:
A:Gene: Lhca2; Cab5
C:Superfamily: chlorophyll a/b-binding protein
C:Keywords: photosystem I; transmembrane protein

```
Query Match      12.7%; Score 84.5; DB 2; Length 271;
Best Local Similarity 35.1%; Pred. No. 0.16;
Matches 20; Conservative 6; Mismatches 20; Indels 11; Gaps 2;
```

RESULT: 6
S52261
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) flavoprotein 1 precursor - potato
N:Alternate names: mitochondrial complex I 51K chain; NADH dehydrogenase (ubiquinone)
C:Species: Solanum tuberosum (potato)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 11-Jun-1999
C:Accession: S52261
R:Grohmann, L.; Thieck, O.
submitted to the EMBL Data Library, January 1995
A:Reference number: S52261
A:Accession: S52261
A:Molecule type: mRNA
A:Residues: 1-487 <GRO>
A:Cross-references: EMBL:X83999; NID:g639833; PIDN:CAA58823.1; PID:g639834

C;Function:
A;Description: catalyzes the reduction of ubiquinone to ubiquinol by NADH
A;Pathway: oxidative phosphorylation
C;Superfamily: NADH dehydrogenase (ubiquinone) chain F; NADH dehydrogenase (ubiquinone)
C;Keywords: 4Fe-4S; flavoprotein; FMN; iron-sulfur protein; membrane-associated complex;
F:92-470/Domain: NADH dehydrogenase (ubiquinone) I chain F homology <NUOP>
F:110-119/Region: NAD binding motif
F:223-238/Region: FMN binding motif
F:402,405,408,448/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 11.8%; Score 78.5; DB 1; Length 487;
Best Local Similarity 27.2%; Pred. No. 1.3;
Matches 25; Conservative 9; Mismatches 33; Indels 25; Gaps 4;

QY 36 PAPFCPLPDPMP-FGALAGD--VLPVYGYHPDFAKIDMRVEWPRS-----GEPWAP 86
DB 41 PQTPTPPPPERTHFGGLKDEIRFTNLXLGHPYLKAMKRGDWYRTKDLVIKGSWDIV 100

QY 87 DPAKSLAGNLGHKDLISFRTPLDGLGGASF 118
DB 101 NEMKK-----SGLRGGGAGF 116

RESULT 7
alpha, alpha-trehalose-phosphate synthase (UDP-forming) (EC 2.4.1.15) 56K chain - yeast
N;Alternate names: Cif1 protein; general glucose sensor GGS1; protein YBR0922; protein X
C;Species: Saccharomyces cerevisiae
C;Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text change 29-Oct-1999
C;Accession: S34979; S31310; S46569; S47858; S45994; S22259; S27249; S36870; S44
R;van Aelst, L.; Hohmann, S.; Bulaya, B.; de Koning, W.; Sterkstra, L.; Neves, M.J.; Luy
P.; Vanhalewyn, M.; Durnez, P.; Jans, A.W.H.; Thevelein, J.M.
Mol. Microbiol. 8, 927-943, 1993
A;Title: Molecular cloning of a gene involved in glucose sensing in the yeast Saccharomy
A;Reference number: S34979; MUID:93360822
A;Accession: S34979
A;Molecule type: DNA
A;Residues: 1-495 <VAN>
A;Cross-references: EMBL:X67499; NID:g3741; PIDN:CAA47834.1; PID:g3742
R;McDougall, J.; Kaasen, I.; Strom, A.R.
submitted to the EMBL Data Library, September 1992
A;Reference number: S31310
A;Accession: S31310
A;Molecule type: DNA
A;Residues: 1-495 <MCD>
A;Cross-references: EMBL:X68496; NID:g4628; PIDN:CAA48510.1; PID:g4629
R;Becam, A.M.; Cullin, C.; Grzybowska, E.; Lacroute, F.; Nasr, F.; Ozier-Kalogeropoulos,
Yeast 10(Suppl.A), S1-S11, 1994
A;Title: The sequence of 29.7kb from the right arm of chromosome II reveals 13 complete
A;Reference number: S46569; MUID:94378717
A;Accession: S46569
A;Molecule type: DNA
A;Residues: 1-495 <BEC>
A;Cross-references: EMBL:X75891; NID:g496856; PIDN:CAA53485.1; PID:g496857
R;Becam, A.M.; Herbert, C.J.; Nasr, F.; Slonimski, P.P.; Zagulski, M.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45995
A;Accession: S45995
A;Molecule type: DNA
A;Residues: 1-495 <BEC>
A;Cross-references: EMBL:X75891; NID:g496856; PIDN:CAA53485.1; PID:g496857
R;Becam, A.M.; Herbert, C.J.; Nasr, F.; Slonimski, P.P.; Zagulski, M.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45995
A;Accession: S45995
A;Molecule type: DNA
A;Residues: 1-495 <BEC>
A;Cross-references: EMBL:X75891; NID:g496856; PIDN:CAA53485.1; PID:g496857
R;Becam, A.M.; Herbert, C.J.; Nasr, F.; Slonimski, P.P.; Zagulski, M.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45995
A;Accession: S45995
A;Molecule type: DNA
A;Residues: 1-495 <CAN>
A;Cross-references: EMBL:L21999; NID:g347489; PIDN:AAA53672.1; PID:g347490
R;Feldmann, H.; Mannhaupt, G.; Schwarzlouse, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45927

A;Accession: S45994
A;Molecule type: DNA
A;Residues: 291-495 <FEL>
A;Cross-references: EMBL:X35995; MIPS:YBR126C
R;Gonzalez, M.I.; Stucka, R.; Blazquez, M.A.; Feldmann, H.; Gancedo, C.
Yeast 8, 183-192, 1992
A;Title: Molecular cloning of Cif1, a yeast gene necessary for growth on glucose.
A;Reference number: S22259; MUID:92245756
A;Accession: S22259
A;Molecule type: DNA
A;Residues: 1-128,'FG',131-495 <GON>
A;Cross-references: EMBL:X61275; NID:g3539; PIDN:CAA43580.1; PID:g3540
R;Bell, W.; Klaassen, P.; Ohnacker, M.; Boller, T.; Herweijer, M.; Schoppink, P.; van
Eur. J. Biochem. 209, 951-959, 1992
A;Title: Characterization of the 56-kDa subunit of yeast trehalose-6-phosphate syntha
tion.
A;Reference number: S27249; MUID:93049353
A;Accession: S27249
A;Molecule type: mRNA
A;Residues: 1-74,'W',76-100,'L',102-103,'W',105-216,'S',218-287,'L',289-495 <BEL>
A;Cross-references: EMBL:X68214; NID:g4626; PIDN:CAA48296.1; PID:g4627
R;Vuorio, O.E.; Kalkinen, N.; Lonsdesborough, J.
Eur. J. Biochem. 216, 849-861, 1993
A;Title: Cloning of two related genes encoding the 56-kDa and 123-kDa subunits of tre
A;Reference number: S36868; MUID:94009048
A;Accession: S36870
A;Molecule type: protein
A;Residues: 63-64,'X',66-71;196-200;223-244;280-283;294-298;304,'X',306-313,'X',315-3
R;Feldmann, H.
submitted to the EMBL Data Library, April 1994
A;Reference number: S44670
A;Accession: S44704
A;Molecule type: DNA
A;Residues: 291-495 <FE2>
A;Cross-references: EMBL:X78993
R;Nelson, H.; Mandiyan, S.; Nelson, N.
J. Biol. Chem. 264, 1775-1778, 1989
A;Title: A conserved gene encoding the 57-kDa subunit of the yeast vacuolar H⁺-ATPase
A;Reference number: A31487; MUID:89109197
A;Accession: D31487
A;Molecule type: DNA
A;Residues: 366-495 <NEL>
A;Cross-references: EMBL:J04450; NID:g171121; PIDN:AAA66891.1; PID:g808824
A;Note: discrepancies are due to frameshift errors
C;Genetics:
A;Gene: SGD:TPS1; Cif1; GGS1; TSS1; GLC6
A;Cross-references: SGD:S0000330; MIPS:YBR126C
A;Map position: 2R
C;Function:
A;Description: catalyzes the formation of alpha,alpha-trehalose 6-phosphate from UDPg
A;Pathway: trehalose biosynthesis
C;Keywords: glycosyltransferase; hexosyltransferase; trehalose biosynthesis

Query Match 10.9%; Score 72.5; DB 2; Length 495;
Best Local Similarity 23.7%; Pred. No. 5.7;
Matches 22; Conservative 19; Mismatches 29; Indels 23; Gaps 4;

QY 14 KDAVEKFPAPLYVCWENHLMFPAPFCPLPPDPMPFGALAGDVLPPVGYGHPDFAKIDWD 73
DB 80 KDLLEKNFNAVPIFLSDE-----IADLHYNGFSNLIWPLFHYHP--GEINFD 124

QY 74 RVEWF---RSGEPWAPDPAKSLAGNGLGHKDLI 103
DB 125 ENAWLAYNEANQTTNEIART-----MNHNDLI 152

RESULT 8
D47301

VlrB9 homolog - Bordetella pertussis
C:Species: Bordetella pertussis
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Nov-1994
C:Accession: D47301
R:Weiss, A.A.; Johnson, F.D.; Burns, D.L.
Proc. Natl. Acad. Sci. U.S.A. 90, 2970-2974, 1993
A:Title: Molecular characterization of an operon required for pertussis toxin secretion.
A:Reference number: A47301; MUID:93219406
A:Contents: BP338
A:Accession: D47301
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-273 <WEI>
A:Note: sequence extracted from NCBI backbone (NCBI:128775, NCBI:128780)

Query Match 10.7%; Score 71.5; DB 2; Length 273;
Best Local Similarity 21.6%; Pred. No. 3.6;
Matches 38; Conservative 20; Mismatches 47; Indels 71; Gaps 8;
QY 3 VIALKPYDFPVKDAVEKFPAPLLYVCWENHLM----- 34
DB 83 VLIKAKSFAQGSQAQPEGL---WNTLLVKTDRRLYDFDLVLASADAATPQALQRS 139
QY 35 -----FP-----APFCLPLPPDPFGAL-----AGDVLPPVYGYHDPFAKI 70
DB 140 RMAYRLQFYRYPAQAASRASPVGPAVAGALNRRYAMQVNGSDGIAPTAAY----- 192
QY 71 DWRVVEW--FRSGEPW-----APDPAKSLAGNGLGHKDLISFR--TPGLDGLGAS 117
DB 193 DGRHRTWLTFRPGQFPFAVFAVADPTGTETLVNLHIDNQSLVIRHAPVLMRLSGAS 248

RESULT 9
S69546
phosphoenolpyruvate carboxykinase (GTP) (EC 4.1.1.32) precursor, mitochondrial - human
C:Species: Homo sapiens (man)
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 22-Jun-1999
C:Accession: S69546
R:Nodaresi, S.; Christ, B.; Bratke, J.; Zahn, S.; Heise, T.; Jungermann, K.
Biochem. J. 315, 807-814, 1996
A:Title: Molecular cloning, sequencing and expression of the cDNA of the mitochondrial
A:Reference number: S69546; MUID:96220691
A:Accession: S69546
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-640 <MOD>
A:Cross-references: EMBL:X92720; NID:g1403049; PIDN:CAA63380.1; PID:e209003; PID:g140305
A:Note: the authors translated the codon GCC for residue 477 as Arg
C:Genetics:
A:Genome: nuclear
C:Superfamily: phosphoenolpyruvate carboxykinase (GTP)
C:Keywords: carbon-carbon lyase; carboxy-lyase; mitochondrion

Query Match 10.7%; Score 71.5; DB 2; Length 640;
Best Local Similarity 25.4%; Pred. No. 9.7;
Matches 34; Conservative 16; Mismatches 51; Indels 33; Gaps 7;
QY 6 LKPYDFPVKDAVEKFPAPLLYVCWENHLMFPAPFCLPLPPDPFGALAGDVLPPVY---- 61
DB 135 MSPADQF-RAVDERFPG-----CMQGRITMYLVLPFSMG-PVGSPLSRIGVQLTDSAYVVAS 187
QY 62 -----GYHDPFAKTDWRVWFRS-----GEP---WAPDPAKSLAGNGLGHKDLI 103
DB 188 MRITWRLGTPVQLAGDGFVLCVLSVGOPLTQGGEPVSWQPCNPKTLIGHVPDQREII 247
QY 104 SFRTPGLDGLGAS 117
DB 248 SFGS-----GYGGS 257
RESULT 10

S52216
viral proteinase - rabies virus
C:Species: rabies virus
C:Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 21-Aug-1998
C:Accession: S52216
R:Camacho, A.; Tabaro, E.
submitted to the EMBL Data Library, June 1994
A:Reference number: S52215
A:Accession: S52216
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1-522 <CAM>
A:Cross-references: EMBL:X79983
A:Note: the source is designated as pseudorabies virus
C:Superfamily: varicella-zoster virus gene 33 protein

Query Match 10.7%; Score 71; DB 2; Length 522;
Best Local Similarity 23.4%; Pred. No. 8.7;
Matches 22; Conservative 10; Mismatches 22; Indels 40; Gaps 3;
QY 32 HLMFPAPFCLPLPPD-MPFGAL-----AGDVLPPVYGYHDPFAKIDWRVWFR 79
DB 349 HVRYPTPTYPAPAPQLPPGAVGVGPHPHPHHAAGALYPPMYAPQGL----- 396
QY 80 SGEPWAPDPAKSLAGNGLGHKDLISFRTPGLDGL 113
DB 397 -----HAPPFSP-----VAHAVPALPGL 414

RESULT 11
S26045
sex-determining protein transformer - fruit fly (Drosophila hydei)
C:Species: Drosophila hydei
C:Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 17-Mar-1999
C:Accession: S26045
R'O'Neill, M.T.; Belote, J.M.
Genetics 131, 113-128, 1992
A:Title: Interspecific comparison of the transformer gene of drosophila reveals an un
A:Reference number: S26044; MUID:92275311
A:Accession: S26045
A:Molecule type: DNA
A:Residues: 1-201 <ONE>
A:Cross-references: EMBL:X66931; NID:g7457; PID:g7458
A:Note: the authors did not translate the codon for residue 111
C:Genetics:
A:Gene: tra
A:Cross-references: FlyBase:FBgn0012404
A:Introns: 11/3; 134/3

Query Match 10.2%; Score 68; DB 2; Length 201;
Best Local Similarity 33.3%; Pred. No. 5.9;
Matches 20; Conservative 7; Mismatches 27; Indels 6; Gaps 3;
QY 35 FPAPFCLPLPPDMP-FGALAGDV---LP--PVYGYHDPFAKIDWRVWFRSGEPWAPDP 88
DB 130 YPAYAWPPPPQAPQPNMYGAVPYGMPSPRYPAHYEAPYPRPLTTPYRAPPRPHP 189

RESULT 12
T40809
conserved hypothetical zinc-finger protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40809
R:Beck, A.; Reinhardt, R.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z21949
A:Accession: T40809
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-482 <BEC>

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:30:09 : Search time 58.11 Seconds
(without alignments)
62.927 Million cell updates/sec

Title: US-09-430-029-6

Perfect score: 666

Sequence: 1 MAVTALKPYDFPVKDAVEKF.....HKDLISFRTPLDGLGGSF 118

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_38.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	215.5	32.4	119	1	DMPO_PSESP
2	127	19.1	670	1	TBUD_BURPI
3	74	11.1	1189	1	HAIR_HUMAN
4	72.5	10.9	495	1	TPS1_YEAST
5	71.5	10.7	286	1	FHHD_ALCEU
6	71.5	10.7	640	1	PPCM_HUMAN
7	68	10.2	201	1	TRSF_DROHY
8	67.5	10.1	405	1	LIM1_BRARE
9	67	10.1	1181	1	HAIR_RAT
10	66.5	10.0	345	1	CEBB_HUMAN
11	66.5	10.0	642	1	PISA_STRAT
12	65	9.8	464	1	NUBM_HUMAN
13	65	9.8	491	1	PRCC_HUMAN
14	64.5	9.7	396	1	DPS1_PINST
15	64.5	9.7	396	1	DP52_PINST
16	64.5	9.7	437	1	MYCN_XENLA
17	64.5	9.7	502	1	VL2_HPV48
18	64	9.6	342	1	P2_ARATH
19	64	9.6	443	1	HXA3_MOUSE
20	64	9.6	464	1	NUBM_BOVIN
21	64	9.6	705	1	C1R_HUMAN
22	64	9.6	1182	1	HAIR_MOUSE
23	63.5	9.5	245	1	Y882_HAEIN
24	63.5	9.5	511	1	CP4B_HUMAN
25	63	9.5	515	1	ENG_BLV
26	63	9.5	529	1	GAG_MLVDU
27	63	9.5	1048	1	SRA4_RAT
28	62.5	9.4	345	1	PL_ARATH
29	62.5	9.4	382	1	LMX1_MESAU
30	62.5	9.4	736	1	DVL2_HUMAN
31	62	9.3	420	1	PHR_THETH
32	61.5	9.2	397	1	ASSY_METVA
33	61.5	9.2	493	1	YLD2_CAEEEL

RESULT 1
DMPO_PSESP
ID DMPO_PSESP STANDARD; PRT; 119 AA.
AC P19733;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE PHENOL HYDROXYLASE P4 PROTEIN (EC 1.14.13.7) (PHENOL 2-MONOOXYGENASE
DE P4 COMPONENT)
GN DMPO OR PHEA5.
OS Pseudomonas sp. (strain CF600).
OG Plasmid pVil150.
OC Bacteria; Proteobacteria.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91072230.
RA Nordlund I., Powlowski J., Shingler V.;
RT "Complete nucleotide sequence and polypeptide analysis of
RT multicomponent phenol hydroxylase from Pseudomonas sp. strain
RT CF600".
RL J. Bacteriol. 172:6826-6833(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BH;
RA Takeo M., Maeda Y., Okada H., Miyama K., Mori K., Ike M.,
RA Fujita M.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CATABOLIZES PHENOL, AND SOME OF ITS METHYLATED
CC DERIVATIVES. P4 IS REQUIRED FOR GROWTH ON PHENOL, AND FOR
CC IN VITRO PHENOL HYDROXYLASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: PHENOL + NADPH + O(2) -> CATECHOL + NADP(+) + H(2)O.
CC -!- COFACTOR: FAD, AND REQUIRES FE(+2) FOR ACTIVITY.
CC -!- PATHWAY: FIRST STEP OF PHENOL BIODEGRADATION.
CC -!- SUBUNIT: THE MULTICOMPONENT ENZYME PHENOL HYDROXYLASE IS FORMED
CC BY P0, P1, P2, P3, P4 AND P5 POLYPEPTIDES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M60276; AAA25943.1; -.
DR EMBL; D28864; BAA06018.1; -.
DR PIR; E37831; E37831.
KW Aromatic hydrocarbons catabolism; Oxidoreductase; Monooxygenase;
KW NADP; Flavoprotein; FAD; Iron; Plasmid.
SQ SEQUENCE 119 AA; 13207 MW; AE0151A918638C49 CRC64;

Query Match 32.4%; Score 215.5; DB 1; Length 119;
Best Local Similarity 37.6%; Pred. No. 2.5e-15;
Matches 44; Conservative 18; Mismatches 54; Indels 1; Gaps 1;

RESULT	3
HAIR_HUMAN	
ID	HAIR_HUMAN STANDARD; PRT: 1189 AA.
AC	O43593;
DT	15-JUL-1999 (Rel. 38, Created)
DT	15-JUL-1999 (Rel. 38, Last sequence update)
DT	15-JUL-1999 (Rel. 38, Last annotation update)
DE	HAIRLESS PROTEIN.
GN	HR.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.
RN	[1]
RP	SEQUENCE FROM N.A., AND VARIANT ALUNC ALA-1022.
RX	MEDLINE; 98111413.
RA	Ahamad W., Ul Haque M.F., Brancolini V., Tsou H.C., Ul Haque S.,
RA	Lam H.H., Alta V.M., Owen J., Deblaquiere M., Frank J.,
RA	Cserhalmai-Friedman P.B., Leask A., McGrath J.A., Peacocke M.,
RA	Ahamad M., Ott J., Cristiano A.M.;
RT	"Alpecia universalis associated with a mutation in the human hairless
RT	gene.";
RL	Science 279:720-724(1998).
RN	[2]
RP	VARIANT APL GLN-620.
RX	MEDLINE; 98431781.
RA	Ahamad W., Irvine A.D., Lam H., Buckley C., Bingham E.A.,
RA	Panteleyev A.A., Ahmad M., McGrath J.A., Christiano A.M.;
RT	"A missense mutation in the zinc-finger domain of the human hairless
RT	gene underlies congenital atrichia in a family of Irish travellers.";
RT	Am. J. Hum. Genet. 63:984-991(1998).
RN	[3]
RP	VARIANT ALUNC ASP-1136, AND ALTERNATIVE SPLICING.
RX	MEDLINE; 98409496.
RA	Cichon S., Anker M., Vogt I.R., Rohleder H., Putzstuck M., Hillmer A.,
RA	Farooq S.A., Al-Dhafri K.S., Ahmad M., Haque S., Rietschel M.,
RA	Propping P., Kruse R., Nothen M.M.;
RT	"Cloning, genomic organization, alternative transcripts and mutational
RT	analysis of the gene responsible for autosomal recessive universal
RT	congenital alopecia";
RL	Hum. Mol. Genet. 7:1671-1679(1998).
CC	-!- FUNCTION: MAY ACT AS A TRANSCRIPTION FACTOR THAT COULD ACT ON TO
CC	REGULATE ONE OF THE PHASES OF HAIR GROWTH.
CC	-!- SUBCELLULAR LOCATION: NUCLEAR.
CC	-!- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS ARE PRODUCED BY
CC	ALTERNATIVE SPLICING.
CC	-!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN AND SKIN.
CC	-!- DISEASE: DEFECTS IN HR ARE THE CAUSE OF ALOPECIA UNIVERSALIS
CC	(ALUNC); A RARE AUTOSOMAL RECESSIVE FORM OF HAIR LOSS. IT IS
CC	CHARACTERIZED BY HAIR FOLLICLES WITHOUT HAIR.
CC	-!- DISEASE: DEFECTS IN HR ARE THE CAUSE OF ATRICHIA WITH PAPULAR
CC	LESIONS (APL) (ALSO KNOWN AS CONGENITAL ATRICHIA) THIS AUTOSOMAL
CC	RECESSIVE DISEASE IS CHARACTERIZED BY PAPILLARY LESIONS OVER MOST
CC	OF THE BODY AND ALMOST COMPLETE ABSENCE OF HAIR.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	the European Bioinformatics Institute. There are no restrictions on
CC	use by non-profit institutions as long as its content is in no
CC	modified and this statement is not removed. Usage by and for commer-
CC	entities requires a license agreement (See http://www.isb-sib.ch/announ-
CC	or send an email to licensee@sb-sib.ch).
CC	-----
DR	EMBL; AF039196; AAC32258.1; -
DR	MIM; 602302; -
DR	MIM; 203655; -
DR	MIM; 209500; -
KW	Zinc-finger; DNA-binding; Nuclear protein; Transcription regulation;
KW	Metal-binding; Alternative splicing; Disease mutation.
FT	ZN_FING 600 625
FT	VARIANT 620 620
FT	/FTid=VAR_005265.